Educational Evaluation: Organization, Bureaucracy and Participation

Evaluation in Norwegian and Finnish primary schools

Henrik Karlstrøm

Thesis submitted as partial fulfillment of the requirement for the degree of Master of Philosophy in Comparative and International Education.

UNIVERSITY OF OSLO

Institute for Educational Research

Faculty of Education

Spring 2008

Abstract

The main purpose of this study is to examine the relationship between evaluation and organizational forms in Norwegian and Finnish primary schools. With a field work in Norwegian schools, an in-depth analysis of this system is provided, with Finland chosen as a comparative example to contrast and compare with the Norwegian system.

The field work was conducted in three primary schools in Norway. About twenty-five teachers, principals and pupils were interviewed in semi-structured qualitative interviews to provide me with enough data to make an analysis. The main questions being examined revolved around the practice of evaluation in the two countries, its connection to the organizational environment of the school systems and the level of participation in the evaluation process among stakeholders both inside and outside schools.

The results show that there is a clear difference in how Norway and Finland conduct and assess evaluation, and that there are advantages and drawbacks to both methods. Norwegian schools are evaluated with the classical approach, and could do with a certain loosening up of their evaluation structure, while Finnish schools have moved towards a stakeholder approach to provide more institutional autonomy, but might need some external guidance to fully utilize their potential.

Acknowledgments

My thanks go to all those in the schools I have visited, as well as those who helped me get into contact with them. The teachers who put aside their all-important work to answer my questions deserve a special thank you, as well as those principals and administrators who arranged my interviews, helped me get the data I needed, and showed me where to go to get coffee.

Many thanks also to my supervisor, professor Arild Tjeldvoll of the Institute of Educational Research, University of Oslo, for professional guidance, useful conversations, and that important first contact with the subject area.

For proofreading and free lunches I thank my mother, Nina Karlstrøm.

Special gratitude goes to Mona, who got me out of bed in the mornings. Without her support there would be no thesis.

Henrik Karlstrøm,

Oslo, April, 2008

Table of Contents

Abbreviations

FGE – Fourth Generation Evaluation

HRM – Human Resource Management

NESH – Nasjonal Forskningsetiske Komité for Samfunnsvitenskap og Humaniora [National Committee for Research Ethics in Social Sciences and the Humanities]

NOU – Norsk Offentlig Utredning [Norwegian Public White Paper]

OECD – Organization of Economic Cooperation and Development

PIRLS – Progress in International Reading Literacy Study

PISA – Programme for International Student Assessment

SFO – Skolefritidsordningen [After School Activity Service]

TIMMS – Trends in International Mathematics and Science Study

A note on translations

The interviews in this thesis were all conducted in Norwegian, and some of the sources cited are in other languages than English. All translations into English are done by me, and any mistakes and imprecisions are entirely my fault. This of course applies to any spelling mistakes and/or bad English as well.

1. Introduction

This thesis is an examination of the process of evaluation in the primary school sectors of two Nordic countries, and how this process is received and negotiated by the stakeholders involved. What does this have to say for the general discussion revolving around evaluation issues, and what can it tell us about the challenges facing Norwegian and Finnish schools?

In this chapter I will lay out some of the main themes of the thesis. I then provide some background to the thesis, describing the context of the two countries with their similarities and disparities. I will also provide a rationale for why this research question is interesting and give a short discussion of the research question itself, along with some underlying assumptions stemming from it. The chapter concludes with a brief summary of the methods used to gather and analyze data and an outline of a general structure of the rest of the thesis.

1.1 Main themes

This thesis concerns itself with the use of evaluations within primary schools in Norway and Finland. The main theme of the thesis is how key stakeholders within the two education systems perceive the process of evaluation and the work that goes into it, as well as how these evaluations affect the institutions being evaluated.

This immediately gives rise to some interesting questions. How is evaluation done? What importance is it given? How do intra-organizational relations affect the reception of the evaluation results? What comes out of evaluation? What is the factual role of evaluation, and how is it perceived? As a tool for improvement, or just another unpleasant bureaucratic duty?

Although each of the questions above could be the subject of a separate thesis, they are all part of an attempt to describe the factual situation in educational institutions in Norway and Finland today. The goal is to tie this description together with some more theoretical considerations around evaluation. These theoretical considerations will deal with questions

of accountability, control, democratic decision making and deliberation and the sociology of institutions.

1.2 Background

In order to understand the situation of the schools in Norway and Finland regarding evaluation, some background is needed. I here present some important factors that help give context to my findings and inform my analysis. The information of general importance is related to current developments in the managerial sector, namely a historically steady rise in the proliferation of evaluation, an increased focus on cross-national tests of learning outcomes, the rise of the managerial system New Public Management and budgetary dilemmas of the modern welfare state. The more local background is related to the specific Nordic welfare states and differences in teacher education and recruitment between Norway and Finland.

1.2.1 Rise in evaluation

The use of evaluation of educational institutions and programs has increased dramatically the last 15-20 years. Neave (1998) calls it “the rise of the Evaluative State”. It is against this backdrop one must understand the discussion about what the evaluation of education is really about.

Evaluation is nothing new. It has always been part of human action. However, like most other human action, it is only during the last century that it has become formalized and subject to formal procedures and methodology. Where it used to be only a set of procedures to be followed internally for those companies that wanted it, evaluation is now big business, and the process of evaluation is often mandatory (Mercer 2005). The field of auditing and quality assurance has gone through a stage of professionalizing and method development (Power 1997), like many other fields in the grey area between practice and theory. It has developed what, on the surface at least, looks like sound scientific methods of assessment. In this field of evaluation professionals and pressure from above, many educational institutions

see little choice but to comply with evaluation demands and follow the recommendations provided.

This also applies to the Norwegian and Finnish education sectors. Boyle and Lemaire (1999) identify two "waves" of evaluation:

> In the "second wave", starting from the end of the 1970s, are other countries which have made significant strides in institutionalizing evaluation, such as Norway, Denmark, the Netherlands, Great Britain, Finland, and France (1999:1)

Municipalities and other school owners routinely commission school evaluations, and during a school year a school might have to undertake up to six or seven different evaluations. This is a fairly new development. Finland got its first center for evaluation of education in 2003 (Lyytinen 2006). In Norway this process has not yet been formalized, though the Directorate of Education provides evaluation services for schools and municipalities throughout the country.

1.2.2 International test results

Increasing international attention has been given to the achievement tests of the Organization for Economic Cooperation and Development (OECD), namely the Programme for International Student Assessment (PISA), the Trends in International Mathematics and Science Study (TIMMS) and the Progress in International Reading and Literacy Study (PIRLS). These studies rank the achievement of students in the same age group in all the countries of the OECD area, along with several countries not counted among the OECD countries.

A striking feature of these tests is the distribution of countries. For all of the tests, the countries scoring best are generally Asian. This has caused some to look to this part of the world for inspiration in educational policy matters. However, the single most successful country in all these rankings is not Asian. It is Finland. This is especially interesting, because there is still some uncertainty as to why this country stands apart from its neighbors. None of the other Nordic countries do as well as Finland. This fact alone makes this a country worth looking into.

One of Finland's neighbors, Norway, scores worst among the Nordic countries in these rankings[1]. Yet Norway spends more money on education per student than Finland[2]. The question here is whether the evaluative practices of the two countries have anything to do with this difference in results.

1.2.3 The Nordic welfare model

One of the most interesting questions is how two countries so alike could produce such strikingly different results. They are both social democracies in the tradition of what is now being called the Nordic model of economic organization. This model is built on a large welfare state, high degree of labor organization and a large public sector. As a result, education in both Finland and Norway is free and primary and lower secondary levels are close to 100% public.

Both countries are highly organized, with a similar parliamentary system, although Finland is not a constitutional monarchy like Norway. They are ethnically homogeneous, with Norway having more immigrants than Finland (9% and 2%, respectively[3]), and contain small national minorities in the north. Schooling is mandatory for the first ten years, and more than 90% of all students continue with upper secondary school, for a total of thirteen years of school. Recently, the portion of tertiary education graduates passed 25% in both countries. Since Finns and Norwegians are highly educated, the production of the countries is often specialized and knowledge intensive. The wage structure is compressed, and income disparities are low.

[1] See for example PISA (2003) for a comparison between OECD countries.

[2] http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-NK-05-018/EN/KS-NK-05-018-EN.PDF cited 20.04.08

[3] All statistics in this chapter are from Statistics Norway and Statistics Finland

1.2.4 The need for accountability and budgetary restraint

Even with large public sectors and high tax revenues in both countries, Norway and Finland are faced with the same dilemmas that face all modern welfare states. As the rights of the citizens to health care, education and other social benefits keep increasing while the acceptance for high taxes goes down[4], the public sector is forced to prioritize within ever more limited budgets. This creates a drive for budgetary restraint and accountability among public service providers: "At all levels, interest in linking budgets to performance measures has resurfaced" (Duncombe, Miner & Ruggiero 1997:1).

One way to ensure accountability is to subject school practice to evaluation on a regular basis. A cornerstone of evaluation is that it guarantees transparency and accountability from those being evaluated, because it makes it possible to assess what is being done right or wrong and place responsibility where it belongs. One of the major themes of this thesis is the link between accountability and evaluation.

1.2.5 New Public Management

This development of ever increasing evaluation has come about as a result of several factors, but perhaps the most important one is the emergence of the system of administration called New Public Management.

According to Karlsen (2002), this development has its roots in four developments, converging towards this system of governance. The first is the mounting challenges to the welfare state model, described above. The second is what Karlsen calls the "crisis of management", a situation where the system of governance has become too decentralized, and the government see a need for a tightening of the structure of governance. The third is the ideological changes in the Western countries, where the earlier social democratic tradition was challenged by a reinvigorated and reinvented economic liberalism. The fourth backdrop to the emergence of New Public Management is the process of increasing

[4] The general level of taxes in Norway and Finland have decreased slightly the last fifteen years, while the purchase power has been increasing with just less than 4% annually in the same period.

international trade and movement of goods, services and labor combined with continuing political integration of regions of the world that is usually referred to as globalization.

This system of administration has come out of the dilemmas mentioned above, where the public sector needs to be restrained as the willingness to pay is disproportionate to the demands made on the system. The New Public Management system takes its clues from private corporation bureaucracies in trying to keep costs down. This often involves cost pricing every transaction and service rendered, to make visible all expenses and make it possible to identify areas where this is not done optimally.

In theory, this should lead to more precise understanding of the expenses in the public system, and hence to more efficiency as wasteful practices are abandoned. However, the system is not without controversy. Firstly, it is feared that the focus on cost efficiency will overshadow the institutions' chief concerns, namely operating as institutions of learning. Secondly, the strict control measures required to keep track of every expense involves a considerable amount of bureaucracy in itself.

1.2.6 Teacher recruitment

One notable difference when comparing the two countries is the teacher recruitment situation. In Norway there is a severe problem with recruiting enough people to the teaching profession. The last years have seen a steady decrease in the number of applicants to the general teacher training. As interest in becoming a teacher wanes, the teacher training institutions have to lower the entrance requirements to have a chance at replacing the large cohort of teachers that will be retiring in the next five to ten years. Possible explanations for this phenomenon are that teacher status is low, or that salaries are low, or both. The general labor market is so favorable to employees in Norway at present[5] that both these factors combine to make it unattractive to apply for a job as teacher in Norway today.

[5] Unemployment in the last quarter of 2007 at 2.1% (Statistics Norway).

This is in stark contrast to Finland. Here recruitment to the teaching profession is good, and has been since World War II. Itkonen & Jahnukainen (2007) state that teachers enjoy a lot of respect, and that the position of teacher is sought after. Finnish school principals report that teacher shortage is among the least likely factors to hinder instructional capacity. Less than 10% of Finnish pupils are affected by teacher shortage, while more than 20% of American pupils are.

1.2.7 **Different teacher education**

Teacher education in Norway and Finland are fairly similar, but there are some differences. There is a high degree of specialization at an earlier stage in Finland. Teaching requires five years higher education, with at least one large specialization. According to Niemi (2006:42), Finnish teacher training has a number of characteristics. The exam has an academic level and most of the class teacher students finish their study. Class teachers have a positive perception of the teacher profession and the convenience of teacher's work tasks. One of the most important aspects of the subject teacher's education is the solid connection between research and subject. The teacher education has high status, as only 10 – 15 % of the people that apply for the class teacher education get accepted. Talented students apply for the education. Young teachers view teachers' work as developing constantly. The students had high-quality subject knowledge and ability to plan teacher lectures.

Norway has four years general teacher's education for teachers at primary and lower secondary school level. This training is more general, and the teachers are expected to teach a wide range of subjects. There is also a possibility to become a teacher after attaining a master's degree at a higher educational institution. This requires at least one year of pedagogical training, and qualifies for work in upper secondary education. Traditionally all students applying for the Norwegian teacher education has been admitted to the education. Further, the students that are admitted are described as been of low quality (Mølstad 2008). The Norwegian government has implemented common educational reforms to increase the capability of competing internationally. In Norway the teacher's role is to facilitate for pupils, not having the traditional teacher role.

1.3 Rationale

Evaluation is important for several reasons. I will go into why this issue deserves further investigation, in order to give a rationale for the thesis. There are three main reasons why evaluation of educational institutions is important. The first is that evaluation practices might influence the way teaching is done, and hence have an impact on learning outcomes. The second is that the work being done before, during and after evaluations can tell us much about the current power situations within organizations such as schools. The third is that the issue of evaluation plays into a more general theme of organizational autonomy, accountability and control.

1.3.1 Evaluation and school practice

The stated goal of many evaluations is to improve current school practice. By helping to identify areas where something can be done more efficient or with better quality, evaluation can serve as a part of the process of improvement. However, it is not given that evaluation will work for improvement.

Evaluation, in order to be of any use at all, has to be part of changing school practice to meet goals that are set by either the organization itself, or the surrounding bureaucracy or political structure. These goals can be anything from improving learning output to employee well-being. Regardless of the stated goal, changing practice in many cases means affecting how teaching, and hence learning, is done in schools. Of course, simply changing practice does not guarantee it is for the better). Thus, one of the rationales for looking at evaluation is its importance for the learning environment of a school, and the impact it can have in the practice of learning.

1.3.2 Intra-organizational power relations

Since the process of evaluation in its very nature involves all parts of an organization, observing the work being done in relation to an evaluation can give us a lot of information on the internal power relations in an organization. The process is constantly negotiated

between the different interests present in a school, both before the evaluation is undertaken, during the process itself and, perhaps most importantly, the work afterwards.

During the planning of education, having the possibility to decide what the subject of evaluation should be, and how it should be conducted, is the source of power within an organization. Similarly, being able to influence the process as it is going on means having the power to influence the final result. The final result is not a given, objective thing. This is where the outcome of the evaluation is decided on, and being in a position to sum up the experience for the whole institution is of great importance.

Different factions within both staff and management will be interested in deciding on the story surrounding the evaluation, and this power struggle can tell us much about how the organization works. Evaluation will therefore have a direct interest for the study of how power is achieved through direct and indirect means within an organizational structure.

1.3.3 Accountability and control

Evaluation is not constricted to individual organizations. It is also a part of a general negotiating of control and autonomy between schools and their surrounding bureaucracies. As accountability and budgetary control is gaining in importance, the process of evaluation plays an ever more important role in justifying expenses or practices. The delegation of evaluation, and the power to decide *how*, *when*, *who* and *what*, is important to identify the potential conflict lines in the public education system.

1.4 Research question

The main research question of the thesis deals with how the procedure of evaluation is conducted in Finnish and Norwegian schools, and the differing perceptions of staff and management within the organizations on the value and use of evaluation. These perceptions are in turn affected by how the stakeholders view the effect of evaluation on quality and the power relations both internally in the organization and externally towards other parts of government. Thus, the question is "*How is the procedure of evaluation conducted in Finnish*

and Norwegian schools, and how do the perceptions of staff and management within the organizations differ on the value and use of evaluation?"

1.4.1 Underlying Assumptions

Even before the investigation starts in earnest, the research question of the thesis implies certain assumptions regarding its theme. In choosing this theme for the thesis, some things are already implied in the formulation of the research question.

There are two major underlying assumptions in my research question, and I will deal with each one here. The first one is that there has been an increase in the use of evaluations of educational programs and institutions in the two countries. Although this could be treated as a quantitative research question of its own, I choose to believe that the general global trend also applies to these two countries. Authors such as Neave (1998) and Power (1997) have described the trend in educational evaluation the last decades and found a steady increase in its use. Official white papers of Norway (NOU 2002:10) and research on Finland (Webb *et al.* 1998) suggest the same.

The second assumption is that the increased use of evaluation has an effect on the quality of education. This should come as no surprise, as the stated goal of most evaluations is to review and, preferably, affect the quality of the program or institution being evaluated. The question is whether the effect is negative or positive. It is not difficult to find reasons why it could be both ways. On the one hand, evaluations are made to give an idea of how a program or institution is going, thus giving an incentive to change into something better. On the other hand, evaluations are themselves processes that move resources and time away from the core business of the institution, and this might affect quality negatively. Also, the process of external evaluation might lead to a feeling of distrust among professionals and funders, something that might also play a part. However, actually assessing the effect of evaluation on quality is outside the scope of this thesis.

1.5 Methodology

The methodological focus of the thesis is a field work in three Norwegian schools, combined with analysis of data from the Finnish education system and a theoretical literature review. The main theoretical discussions revolve around issues of bureaucracy, democratic participation and the conduct of evaluation, specifically focusing on two different approaches to evaluation that I have called the classical approach and the stakeholder approach.

To collect my data, I have chosen to do qualitative interviews in three Norwegian schools. The interviews consist of a set of questions concerning teachers and management and their perceptions of the process of evaluation. This data has then been analyzed using an adapted form of grounded theory coding.

In addition, a review of literature concerning the state of Finnish evaluation is used to paint a picture of contrast and comparison. This review focuses on the differences between Norwegian and Finnish schools regarding teacher and school autonomy and experiences with a different evaluation approach.

1.6 Thesis structure

The thesis is divided into six chapters. The first chapter is this introduction. The second chapter is dedicated to a discussion of theory relevant to the theme of the thesis, among it theories of communicative action, bureaucratic control and participative evaluation. The third chapter is an overview of the methodological choices made in the collection of the data for the thesis. The fourth chapter is the presentation of my findings, along with some first impressions of the data, mostly concentrated on the data collected in the field work. In the fifth chapter I discuss the findings in light of the theoretical perspectives presented in the second chapter, trying to synthesize the empirical findings with the theoretical perspectives. Then follows a small concluding chapter, with a summary of the main points of the thesis and some implications discussed.

2. Theory

In this chapter I describe the theoretical perspectives that inform the analysis of my data. The main point is to give an account of the different approaches to evaluation, and to identify the relevant level at which to approach it.

My theoretical analysis is focused on three aspects of evaluation. The first regards the general philosophical question of evaluation, where issues such as democratic participation, autonomy and control are central. The second has to do with the analysis of the context evaluation is done within today, and is concerned with questions of accountability and trust. The third aspect is theory concerning the practice of evaluation itself: how it is done, what is best practice (if there is one), and evaluation of the evaluation. These three levels are all present in any evaluation, either as a backdrop or in the actual process.

As my research question is "*How is the procedure of evaluation conducted in Finnish and Norwegian schools, and how do the perceptions of staff and management within the organizations differ on the value and use of evaluation?"*, the relevant theory should focus on how evaluation is conducted and the many ways it can be done, along with some reflections on what the intra-organizational structure of power reveals about perceptions of evaluation.

2.1 Philosophy of evaluation

This level of theory is concerned with the philosophical implications of evaluation. How does evaluation tie in with questions of democratic participation, the overlap between professional autonomy and the public sphere, and issues of bureaucratic control? Central to the discussion will be the theories of communicative action and democratic participation put forth by the German sociologist Jürgen Habermas, the theory of bureaucracy advocated by another German sociologist, Max Weber, and the theory of systems of control introduced by the French sociologist Michel Foucault.

2.1.1 Bureaucracy

Evaluation, in the formalized version discussed in this thesis, usually takes place within a formalized bureaucratic framework. Bureaucracies are usually organized into a hierarchical structure that is formally defined, going from the top national/federal level and moving through regional and municipal levels and culminating in the local level. This framework can have differing influence on the outcome and organization of evaluations, and should therefore be discussed in relation to how evaluations are conducted.

The process of evaluation can not be separated from the configuration of national, regional and local bureaucracies. These three levels of bureaucracy have differing, sometimes even conflicting, evaluation demands, and in the always changing configuration of these the current emphasis on evaluation can be found.

One feature of bureaucracies, and perhaps the most important, is their rather permanent nature. A bureaucracy is a form of organization separate from those who inhabit it at any given point in time. In fact, “it takes on rule-like, social fact quality, and when it is embedded in a formal structure, its existence is not tied to a particular actor or situation” (Aldrich 1992).

This relative permanence of bureaucracies is the source of both their strengths and weaknesses. On the one hand, it ensures stability and a certain sense of objectivity. On the other hand, it can lead to rigidity. There is also a problem of convergence of bureaucratic policies: “Once a set of organizations emerges as a field, a paradox arises: rational actors make their organizations increasingly similar as they seek to change them” (DiMaggio & Powell 1983:264).

DiMaggio & Powell argue that the corporate business world has already been through a thorough process of bureaucratization, and that the current development is to turn the process on the state.

> Today, however, structural change in organizations seems less and less driven by competition or by the need for efficiency. [...] bureaucratization and other forms of organizational change occur as the result of processes that make organizations more similar without necessarily making them more efficient (DiMaggio & Powell 1983:265).

This convergence towards homogeneous systems can potentially be problematic. However, the most problematic possibility is the autonomy from the political field. Since politicians come and go while the bureaucracy remains, they often turn into more conservative entities, going their own ways instead of heeding the orders from the legislators.

Much research on bureaucracy is of the opinion that bureaucratic systems tend to converge towards rigid control from the top down. As officials in the central bureaucracy look at educational institutions, they wish for them to adhere to the same rules and regulations as the general state bureaucracy. It might not even be the stated intention of the upper echelons of the bureaucracy to attempt to control the lower levels. There is still a subtle, yet forceful pressure to try to control the outcome of others' work. Max Weber (1990) described this effect, sometimes called the “iron cage” of bureaucracy (or rationality), in his extensive theory of bureaucracy.

However, such a complex entity as a bureaucracy can not be reduced to simple formulae. Many react to the reduction of bureaucracy to a rigid system of control, instead emphasizing the room for negotiation and creativity within the structure (du Gay 2000). The criticism of the anti-bureaucratic movement is that it idealizes one of two states of organization. One ideal way of conducting organization is through “charismatic managerialism”, where what Weber calls charismatic leadership of the state, in which a leader gains authority by virtue of his or hers charismatic powers, is brought into the context of individual businesses or institutions (Weber 1990). The other is through “contemporary communitarianism” (du Gay 2000), with a belief in some version of the Greek proto-democracy transposed onto modern society, where the open forum of the citizens convening to voice their concerns is the model of the ultimate democracy. As du Gay puts it,

> A characteristic feature of the anti-bureaucratic discourse of both charismatic managerialism and contemporary communitarianism is the belied that modern bureaucratic culture signifies the fragmentation of what was, and ideally should again be, a unified civic moral domain. Whether maximum businessing or the reactivated polis is the chosen means to closing the 'wound' that bureaucracy opened is neither here nor there. What the various anti-bureaucrats share is a demand that the 'total pattern of life be made subject to an order that is significant and meaningful' (du Gay 2000:74).

Although the above description of the "anti-bureaucrats" is somewhat overblown, the point has some merit. In a complex and highly organized society, it is hard to envision the organization of all the components without some sort of bureaucratic institutions.

Turning to the more positive description of bureaucracy, the actual responsibility of bureaucracy is often highlighted: "There can be no doubt that state bureaucrats bear a real responsibility for the efficient and economic use and deployment of the resources at their disposal" (du Gay 2000:143). According to the author, bureaucracies manage public funds surprisingly well in most developed economies today. Weber (1990) himself points to the professional "honor" of bureaucrats, meaning that professional bureaucrats, with a certain respect for their public mission, is the only way to guard against corruption and inequitable treatment of those who are in contact with the bureaucracy.

With these reservations in mind, it is possible to discuss the role of bureaucracies in the context of evaluation. The main tangent between evaluation and bureaucracy is in the formulation of specific procedures of evaluation. As the bureaucracy in theory is created to oversee the execution of the policies dictated from the legislative branch of government, it is often their task to transform the rough formulations of law into operational categories.

Depending on the level of bureaucracy being examined, different ways of negotiating the actions of the bureaucracy are enacted in relation to evaluation. At the topmost level, the Ministry of Education is charged with interpreting the laws and regulations passed in the political organs and transforming them into workable instructions. Some of the interpretation is delegated to the levels below, the regional or municipal levels, and the execution is left to the lowest, most local level in the bureaucracy. In any of these levels, the original intent of the lawmakers can be carried through or distorted, depending on the surrounding context.

Together with the professionalization and bureaucratization of evaluative practices has come the inclusion of these into a legal framework. An important factor in the expansion of the public bureaucracy has been the expanding legal framework designed to deal with and give guidelines concerning ever more aspects of life, both privately and professionally. For the specific field of education, this means regulating by law such things as access, curricular content, internal governance and participation in institutional democracy. In Neave's words, "the recourse to legislative enactment as a means of enforcing practice and implementing policy" (1998:269).

The tendency for increasing juridification reflects a more general trend, namely that more and more parts of society are being regulated by laws and regulations. This is not in itself a bad thing, of course, but points back to the bureaucratization of education, which can have some unfortunate side effects.

2.1.2 Control

As the field of educational evaluation grows, it develops a language to describe and justify what is being done. Most of this language is not necessarily constructed for that purpose, but gathered from other disciplines (Power 1997): Some from financial accounting, some from the field of education, some from pedagogy, some from the social sciences and so on. Together they constitute a field of knowledge required to master the field of evaluation, and thus a "technology of power" (Foucault 1980).

The French social theorist Michel Foucault has analyzed professional discourses, and describes how command of these allows control over how a profession is presented, and its professional uses for the "outside world". The point is that, as the use of evaluation both gets professionalized and expanded, the programs and institutions that formerly were not subject to this kind of control now are included in a field of power where they do not command the use of discourse. These forms of disciplinary power are seldom explicitly stated and have no physical manifestation, as opposed to earlier centuries, but do work to internalize discipline in those subjected to it.

According to Derek Layder,

> The individual's own self-monitoring is absorbed as part of the general system of surveillance. This is exemplified in the use of dossiers, marking and classification systems (and other forms of appraisal and monitoring) in schools, hospitals, prisons as well as factories […] society thus has available a means of control – a "technology of power" – that can be deployed at many locations (Layder 1994:100).

Not all evaluation works through hidden agendas and indirect, internalized methods of control. Some policies, like the educational policies of the United States, actively promote the use of evaluations and rankings as a method to "shame" schools into doing better (Carnoy 1999). Also, the institutional framework itself might work as control, even if it is not internalized by those subjected to control or openly stated. Although the following quote is about Human Resource Management, it fits well within the context of evaluation. Simply substitute evaluation for HRM:

> Underlying most studies of HRM, although often remaining implicit, is what may be identified as a systems maintenance or functionalist perspective. Reflecting concerns with improvement in efficiency that derive from classical management theory, HRM is an organizational mechanism through which goal achievement and survival may be promoted. Its aim is to make the organization more orderly and integrated. In HRM, connotations of goal-directed activity, inputs and outputs, stability, adaptability, and systems maintenance predominate (Townley 1993:518).

There is a danger of overstating the effect of control. Not only can there be a legitimate need for control in society, but it is not necessarily something that is felt as any form of oppression. There can be no social relation without forms of power and control being exerted, and so the existence of it in any given social or institutional setting might not signify a relation of dominance. According to Lianos,

> Control is [...] conceived of in terms of arbitrarily presumed restrictive effects and not in terms of a reliable analysis of its production, content, reception and articulation with other social registers (Lianos 2003:414).

This does not mean that questions of control are uninteresting:

> On the other hand, it is necessary to examine the question of social control in relation to the institution, that is to say, the instrument for the conscious and planned management of socialized human activity. In the first place, it is for several reasons very useful to distinguish between control generated by the skein of links between groups or individuals and control deriving from the activity of institutions (Lianos 2003:415).

Lianos goes on to differentiate between the intra-individual type of control, and its institutional effects. The main points are that institutional control is exactly that, institutionalized. “It is produced as a planned *managerial* activity corresponding to the complex mode of organization of contemporary Western society” (Lianos 2003: 415). Control is also integral to certain bureaucratic activities, and is often impossible to separate from processes that are wanted and useful.

Although the increased use of evaluation can be seen as an increase in control, it is not true that those working in education feel like they are under the thumb of some all-powerful *panopticon*. However, it *is* true that an increase in forms of evaluation can mean an increase of control. This might still be exactly what is wanted by society in general. Educational institutions do not operate in a vacuum, and society has its demands.

2.1.3 The classical approach to evaluation

There is no single way to conduct an evaluation, and evaluation has traditionally been done in a wide variety of ways. In fact, House (1981) identifies eight different methods of evaluation, all with different underlying assumptions. However, he sees them all as grounded in the same liberalist ideology of rational and goal oriented behavior.

The classical evaluation models share several traits. Firstly, they are based on individual choice, where the individual is taken to be the source of meaning, and the individual's existence is taken as given before the existence of a society. According to House, the belief in freedom of choice is the singularly most important factor in all the classical models of evaluation. Secondly, they are predominately empirically oriented, often “radically empirical”: “In its most extreme form, the objectivist epistemology completely rules out the non-quantitative” (House 1981:319). Thirdly, they assume a “marketplace of ideas”, where ideas are traded like commodities. Ideas compete, and the best ideas “win”. This way, the optimal strategy is found on the basis of its objective strengths. The consequence of these common traits is in House's view that classical evaluation rules out the societal character of evaluation procedures. This constitutes a democratic problem.

However, House's view is not the only one. While it can be argued that the field of educational evaluation has been dominated by the classical approach, this has not produced singularly negative effects. Reynolds & Teddlie (2001) sum up the many positive results stemming from classical evaluation in the form of school effectiveness studies. The focus on school effectiveness is a typical example of reliance on classical evaluation, with its base in positivist knowledge production.

One striking feature of the classical educational evaluation models is that it has grown from nothing into a mature discipline over the course of few years. It is a fairly young discipline, even if its roots are in much older philosophies and disciplines. Over the course of the "invention" of these models, they have contributed to the generation of a wide knowledge base on diverse fields of education.

Reynolds & Teddlie also argue that the large impact the classical approach has had is a positive thing, in that it has contributed to educational change to a large degree:

> We have convincingly helped to destroy the belief that schools can do nothing to change the society around them, and have also helped to destroy the myth that the influence of family background is so strong on children's development that children are unable to be affected by school (Reynolds & Teddlie 2001:103)

This has shifted the focus away from background factors and over to addressing teacher or system failure. In this way, the focus is on identifying the need for educational change more than the passive acceptance of outside factors as determining children's future.

Although this approach has indeed contributed to the growth of a discipline, and knowledge has been produced, it has met with some criticism. Both the practice and the philosophy have been under fire from proponents of a different procedure. Although Reynolds & Teddlie dismiss the criticisms as "non-rational spasms", it is still interesting to see what they consist of.

2.1.4 The critique of classical evaluation

Summing up the extensive critique of non-stakeholder evaluation is not done quickly. However, according to Weiss (1986), one of the proponents of the stakeholder approach, other forms of evaluation share some problems that are connected to the discussion of participatory democracy and the public sphere. The problems discussed here are tied directly to the discussion of participation in evaluation processes.

According to Weiss, evaluations are by necessity narrowly focused, as it is impossible to include all aspects of the dealings of any organization. However, she claims that "evaluators too often select for attention the issues that are easy to study with available social-scientific tools, not the issues that are important" (Weiss 1986:146). Such a selection practice ignores the issues that it is possible for those involved in a program or institution to change.

Correspondingly, much evaluation data is irrelevant to the practical everyday practice of those involved in education. Outcome evaluation gives little information about *how* improvement should be done, and is thus of little relevance to agents. This again leads to many evaluations being left unused. Evaluation results rarely influence decisions about improving practice. According to Weiss

> The evaluator conducts the study, completes the report, and leaves. Program managers [or institutional leaders] take comfort from the findings that are positive and bury or forget the findings that suggest a need for major reform. [...] Despite all the rhetoric about the utility of evaluative evidence for improving the rationality of decision making, evaluation often seems to leave the situation unchanged (1986:147)

One of the main reasons for these problems is the lack of sensitivity to local needs. Where evaluators were brought in to work at the more local levels, issues specific to the localities were more likely to be heeded. This does not necessarily mean that the concerns of those lower in the hierarchy were addressed, simply that evaluators at the lower levels had more motivation to get involved in concerns below the federal level.

2.1.5 Communicative action and democratic participation

The question of evaluation is intimately tied to participation. Several parties have stakeholder interests in what is going on in educational institutions, including parents, teachers, management and the education bureaucracy. The central questions of any evaluation is therefore who gets to commission it, who gets to be part of it, and who gets to process the resulting information (evaluating the evaluation, so to speak).

According to Habermas (Goode 2005), most of human interaction is guided by communication. The mode of action is most often decided upon through active communication between two or more agents, and this itself is a form of action which Habermas dubs communicative action. The point of this concept is that the communication between agents is the basis of participation in democratic procedures. Thus, communicative action lies at the core of any process involving more than one agent.

Habermas goes on to use the theory of communicative action as the basis for his general theory of democracy in the context of the complex modern nation state. As any action is influenced by the communicative consensus arrived at beforehand, anyone with any form of stake in the outcome of the action should be part of the communicative effort preceding it.

This theory has met some criticism for ignoring the issue of power (Turner 1988). Even though action is decided upon through communication between agents, there exists a problem if those engaging in communicative action are in very different power situations. The relative bargaining position in any communicative situation where a consensus over action is to be reached can have a large influence on the outcome. The idea of a public sphere, where agents communicate to achieve consensus, is more of an “ideal speech situation”, where communication is inherently rational (Brand 1990).

However, Habermas himself tries to divide action into two, where one of them is the communicative actions discussed here, and the other “strategic action”, which is informed by an instrumental purpose in which a person persuades another by “sanctions or gratifications, force or money” (Habermas 1982:269). This type of action is motivated by practical concerns, whereas communicative action is more discursive in nature. It is in communication

arguments that appeals to a common understanding of the situation are made, and where it is possible to shift opinions.

Agreeing with Habermas, Goode is of the view that

> Habermas' analysis does acknowledge the materiality, and not merely the ideology, of the public sphere: unequal patterns of access to time, space, literary skills and the like underpin unequal opportunities to participate in the public sphere. [...] it is also true that, under conditions of increased technological mediation, these problems of material inequality are magnified (Goode 2005:38)

The last point is interesting, as it points to the possible effects of introducing new technology into the public discourse, something that can make a difference in the age of electronic questionnaires.

2.1.6 Stakeholder evaluation

The theory of communicative action has been picked up in the context of evaluation, where it has come to influence the way evaluation procedures is viewed. The main idea is that evaluation is a process with several stakeholders, and therefore those being subjected to evaluation should have a say in the evaluation process. This view is shared by action researchers (Bryman 2004), who seek to empower people working in any vocation so they, in cooperation with the external researcher, can perform research activity within their own social context.

When those affected by evaluation are directly involved in conducting it, it should in theory constitute a more democratic form of evaluation, as this should prevent action being taken over the heads of those involved. The democratic or stakeholder evaluation movement has grown out of criticism of other forms of evaluation that according to the proponents share several weaknesses.

It is with these concerns in mind that the stakeholder approach to evaluation was worked out. The main goal was to “increase the use of evaluations for decision making and to bring a wider variety of people into active participation in the evaluation process” (Weiss

1986:150). The use of stakeholder evaluation will in theory increase fairness in the process as well as empower those affected by it. The net result should be democratization of the evaluation process.

However, several issues remain, even with the stated intent to improve on other forms of evaluation. Among them is the definition of a stakeholder itself. Is it those who make decisions about a process, or all those who are affected by it? Here, the ideas of Habermas come in, as he would define the last group as obviously being a part of what constitutes a stakeholder. Similarly, the American sociologist Nancy Fraser argues that any form of "public sphere" debate going on must also include what she calls "communities of risk", the set of stakeholders affected by developments (Fraser 2005).

More serious is the problem of making stakeholder evaluation compatible with traditional evaluation procedures. Weiss puts it like this:

> The stakeholder approach changes the role of evaluators. They are asked not only to be technical experts who do competent research. They are required to become political managers who orchestrate the involvement of diverse interest groups. They must be negotiators, weighing one set of information requests against others and coming to amicable agreements about priorities. They must be skillful educators, sharing their knowledge about appropriate expectations for program [or institutional] development and program [or institutional] success while giving participants a sense of ownership of the study (Weiss 1986:153).

The question is whether this is too much to ask of evaluators, and whether it is at all realistic to see it happen in real life evaluations. It is also difficult to see traditional evaluators operating in this manner. Evaluation can be complicated business, and it might require extraordinary pedagogical skills to enable participants to have an actual say in the process.

However, examples of this kind of knowledge transfer exist (Yin, Kaftarian & Jacobs 1995)[6]. A project to evaluate the local levels of the federal Centre for Substance Abuse Prevention,

> focused on a "transfer" type of relationship between external evaluation guidance and program participants. In this configuration, the empowering relationship is bidirectional. [...] external evaluators can serve program participants through training, facilitation, and advocacy roles (Yin, Kaftarian & Jacobs 1995:191).

However, it is still unclear whether the stakeholder approach can serve as a universal model for evaluation work, or whether it should be contained to special circumstances.

2.2 Evaluation in context

While the philosophy of education is interesting in and of itself, it can be useful to tie the discussion to the specific context of today. As the education systems of the world are examined more closely in a comparative perspective, the convergence of trends is a logical outcome. These days, the general trend must be examined within the context of accountability for public spending, issues of trust in the education system and New Public Management.

It is the view of several authors that education has become entrenched in policies that promote the running of public institutions as private corporations (Trow 1996; Power 1997; Neave 1998 & Hoppers 2006). The term New Public Management is not universally used for the introduction of market mechanisms in education, but it will be used here as a kind of catch-all term to refer to these developments.

The increased use of New Public Management of public institutions has several potential consequences. The focus of this system of management is on efficiency, public private

[6] The authors call it "empowerment evaluation", but it is nearly indistinguishable from the stakeholder approach described above.

cooperation, accountability and budgetary restraints. The main effects of this focus in the context of evaluation are discussed here.

2.2.1 Efficiency concerns

One concern of policy makers and administrators of education is the efficiency of educational institutions. Is the public getting all it can out of its schools and universities? Could money and resources be put to better uses? As demands made of educational institutions rise while the willingness to fund them through taxes remains stable or even decreases, governments look to other measures, and "seek ways to lessen the amount of resources devoted to [...] education without damaging, if possible, its effectiveness" (Schmidtlein 2004:267). Resource management can be seen as one of the factors behind the increasing use of evaluation in education.

Of course, this kind of direct resource auditing only works if the money saved outweighs the increased cost of constant evaluation. However, even if it does save money, it is still an open question if it is a good idea to go through with, when the additional workload and irritation to the one being submitted to evaluation is factored in.

2.2.2 Market exposure

Another aspect of New Public Management is its openness to market mechanisms. Whereas market exposure would have provoked outcry only a few decades ago, it is now perfectly acceptable to look to markets for ideas and practices. Trow (1996:310) states that "an element of market links can be found in most American institutions", and goes on to say that this is becoming the case in European education as well, although still much weaker.

One interesting side effect of the marketization of education is that educational institutions are expected to work within the market, competing for pupils and staff, while not being entirely trusted to do so in an orderly fashion. As Schmidtlein puts it:

> Paradoxically, implementing this view of institutions as "market place" enterprises often is sought through establishing government administered assessment and accountability processes intended to orient institutions to this "new reality". The new decentralized "industry" paradigm often is coupled with hierarchal, bureaucratic quality assessment management structures (2004:266).

Another trend is the use of quantitative data as basis for decisions. Governments are increasingly engaging in databased quality assessment processes (Schmidtlein 2004). The implementation of these processes is, however, not without its problems. Firstly, it seems to assume that faculty and administrators know how to improve quality but for some reason fail to do so. Secondly, it implies that government officials can assure the public interest in quality through quantitative measuring. Lastly, and perhaps most basic, it assumes that measures of quality can be identified and agreed upon, and that improving quality requires strong bureaucratic coordination and control.

So, is the education sector an industry or not? Should it belong in an education market, competing for pupils and funding? The answers to these questions depend on the real effects of the new developments.

Having established that there has been a move towards new forms of management of public educational institutions in the last twenty years, one has to ask what this means for education. Is it good or bad? What are the changes in the relation between the education sector and society in general?

2.2.3 Educational institutions and society

Educational institutions have a complex relationship with society. They exist within a social and legal structure, and must relate to this. On the one hand, they are society's providers of education. They provide a certain degree of skill and knowledge for the pupils enrolling, and prepare pupils for further studies or work. They have a commitment to the outside society, for providing competent workers to industry and students to higher educational institutions. The balance between knowledge production and knowledge transfer is difficult.

The second commitment of educational institutions is towards the pupils. Education is supposed to make them fit to enter working life with skills that are in demand from society. To do this, the institutions must provide facilities like reading rooms, computer labs and libraries. Even more important is the provision of high quality teaching. It is through instruction and interaction with the teaching staff that pupils learn the skills and knowledge they and society need. Teaching, however, takes time, and is only one of the components of a teacher's workday. What happens when one component suddenly increases, like we have seen in the case of the time allocated to evaluations and report writing?

2.2.4 Assumptions regarding education

According to Frank Schmidtlein (2004), the present interest in governmental assessment and accountability practices is a result of several factors that together signify a new conception of the purposes of educational institutions[7].

The first factor, according to Schmidtlein, is that we are seeing an emerging view of education as an industry rather than as a social institution. According to this view, decreasing public funding, international competition and general market pressures necessitate a move to more business-like models of education management. This includes running educational institutions like commercial entities and seeing them as producers of a product, knowledge, being sold to the consumers, pupils.

The idea of educational institutions as production units fits nicely together with the second factor, concerning doubts about the efficient allocation of resources and effective cost containment (Schmidtlein 2004). The question of spending resources in the right way is crucial to educational institutions, because they are to a high degree “labor intensive”. As schools become more dependent on revenues that fluctuate on a yearly basis, the hunt for increased cost effectiveness intensifies. As Schmidtlein puts it, “this search for a politically

[7] Even if Schmidtlein is predominately interested in higher education, his points are valid across the entire educational sector.

acceptable way to cut budgets for popular programmes undoubtedly is one motivation behind quality assessment" (2004:268).

A third factor is that of the erosion of trust between the official and institutional segments. Martin Trow points out the inevitable trade-off between accountability and trust in the system: "accountability is an alternative to trust; and efforts to strengthen it usually involve parallel efforts to weaken trust" (Trow 1996:311). As the general public loses trust in institutions, measures to check what they are doing with society's resources become more acceptable to implement. This has two effects: it further weakens trust in the institutions, and aggravates sector professionals, who must now spend valuable time, energy and resources to report on their doings.

2.2.5 Trust

In order to function properly, educational institutions require both resources and trust from the general society. Staff, facilities, administration and the running of institutions all cost money, and this money comes from the public or from the pupils. The funders of education, the public or the parents with children in private schools, would like to see their money spent the best possible way. At the same time, the autonomy of educational institutions is universally held in high regard. So the question becomes one of trust versus accountability: Is it possible to demand accountability from schools for how they spend their money, while at the same time leaving room for letting teachers and education managers to take care of the professional side of education?

According to Martin Trow (1996), education is linked to the general society in three ways. One, the market, has been described already. The other two are the concept of trust and the concept of accountability. These two exist in an uneasy relationship with each other. Accountability entails the obligation to report to others and justify the use of resources, and to what effect, while trust is "the provision of support, by either public or private bodies, without the requirement that the institution either provide specific goods and services in return for that support, or account specifically and in detail for the use of those funds" (Trow 1996:310).

These two definitions seem to be mutually exclusive. On the one hand, educational institutions have traditionally enjoyed a lot of autonomy, and still do. They were respected and trusted to deal with their own, while the gains to society were taken at face value. There was no question that society benefited from their existence. On the other hand, these institutions are starting to use a lot of money, and with the spirit of budget restraints gaining ground, one might want to look closer at these institutions. Could we benefit *more* if things were done differently? Accordingly, "accountability to outsiders weakens the autonomy of institutions" (Trow 1996:312). Or, in other words, "there is a tension between governments' legitimate interests in institutional accountability and quality and the values represented by institutional autonomy that have been described by many scholars and practitioners" (Schmidtlein 2004:264).

2.2.6 Accountability

The issue of accountability requires closer scrutiny. There is a tendency to equate quality and accountability, in that an accountable institution performs qualitatively better than an unchecked institution. This is not necessarily false, for how can we *know* if there is no accountability measure? However, there needs to be clarity around which rationales are underlying the accountability measures being taken. According to Lee Harvey and Jethro Newton,

> The lack of clarity arises because the rationale is rarely openly admitted. The rhetoric and documentary preambles in many countries refer to quality evaluation as a process of improvement, yet all the emphases are on accountability, compliance and, in some cases, control of the sector. (Harvey & Newton 2004:151)

Why is there a discrepancy between the underlying rationale, and the rhetoric on the surface? One reason is probably that it is easier to justify evaluations if they refer to some kind of improvement mission, instead of control being the open goal. While everyone can always improve, most people do not enjoy being controlled, or even seen as someone who could and should be controlled. As Carol Weiss puts it,

> Of course, many people do want to know what evaluation can tell them, and many more can be persuaded that evaluation can tell them things that they want to know. But it is equally true that people decide to have a program evaluated for many different reasons, from the eminently rational to the patently political. (Weiss 1998:22)

So, while there is no doubt that a certain degree of accountability in institutions is necessary, this does not mean that all policies initiated for the sake of accountability are for the good.

2.2.7 Evaluation and resources

A lot of evaluation has as purpose, stated or unstated, to improve the use of resources in the institutions and programs being evaluated. Managing scarce resources is a difficult affair, and there are always doubts about whether they are used to maximum effect. However, one must not forget that evaluations take up resources too. Sometimes, they are very costly. Hiring and paying for evaluation staff, administrative costs, any adjustments that have to be made during the course of the evaluation and any changes being made as a result of the evaluation may well end up costing more than the initial money saved by the new practice.

Neave (1998) also points to the fundamental instrumentality in the system of New Public Management, inhibiting academic productivity for the benefit of an increasing management sector. Through the setting of objectives, goals and targets, new managerial procedures forces the education sector to spend more resources on evaluation and self control than before, maybe even outweighing the resources originally saved by introducing the new regime. In the flowery language of Neave, this move "goes hand in hand with a veritable orgy of procedures, audits and elaboration of instruments of administrative intelligence" (Neave 1998:266).

2.3 The practice of evaluation

The above discussion is far removed from the actual practice of evaluation. Yet all work on evaluation must take into account how it is undertaken. As so much evaluation theory is concerned with the method of evaluation, relating to participation, measurements and democracy, it is of crucial importance that the very minute details of evaluation also are

subject to discussion. In this part, I describe some views on *how* evaluation is and should be done, at the detailed local level.

The most important kinds of evaluation are those regarding classical external evaluation and the more action research oriented stakeholder approaches. There is still disparate opinions on which of the two is correct, and so both will be presented here.

2.3.1 The classical approach

The classical approach to evaluation recognizes that evaluation is a professional field, requiring some training to undertake correctly. It usually involves an external evaluation, often, but not necessarily, commissioned from outside the institution being evaluated. It is the responsibility of the commissioner to find someone with the required skills for the evaluation, and who is sufficiently independent from the institution to be impartial in the evaluation.

The point of independence is important, as it is necessary to keep a certain amount of scientific rigor in evaluation, to ensure that the information produced is of the highest quality. Accordingly, evaluation is external, and the evaluator is brought in from the outside. Evaluation is a profession, with its own standards of rigor, and its own scientific methodology.

The process of evaluation will follow a fixed structure, which will often take on a circular quality. The concept, goals and procedure definition and much of the work is done before the evaluation starts. After questions are asked and information gathered, the evaluator produces a report with an assessment of goal achievement and some recommendations for improvement. Those in charge are left to implement change based on the recommendations. A new evaluation is undertaken to oversee whether change has been implemented, and if the changes have had the desired effect.

In order to understand the reasoning behind classical evaluation, it is necessary to go deeper into the methodological issues concerning the classical approach to evaluation. The

somewhat rigid structure is chosen in order to ensure the possibility of continued evaluation. When evaluations measure more or less the same, and do so in the same way, the opportunity to compare over time arises. Also, the work becomes more separated from the person evaluating, allowing the institution being examined more room to solicit extra evaluation and second opinions on evaluation results. A clear definition of concepts and procedure is also beneficial in light of a discussion of the results, as this allows comparison against a concept defined in advance.

Evaluation may very well have control as a stated goal, as there are often legitimate claims to control institutional spending or resource allocation. Thus, an evaluation will often have as a goal to see whether a part of the institutional structure is performing according to the framework it works inside. This can serve an important function in society, allowing for more transparency and democratic control over public institutions.

2.3.2 The stakeholder approach

According to the stakeholder approach, evaluation does not necessarily mean control. Even though the usual reference to improvement in justifications of evaluation is really a front for the real purpose, control, it *can* serve as a tool for improvement. It frequently does, and most often is a mix of those two. However, how do we make sure that an evaluation is for improvement, and not imprisonment?

Tied closely to the stakeholder approach to evaluation is the notion of *empowerment*. The idea is that as practitioners are allowed to take part in the evaluation, they gain ownership to the process. This has obvious democratic potential:

> The effort to devise *joint*, *collaborative*, or *shared* constructions solicits and honours the inputs from the many stakeholders and affords them a measure of control over the nature of the evaluation activity. It is thereby both educative and empowering (Guba & Lincoln 1989:184)

The main methodological approach is to include and solicit help and input from whoever is deemed a stakeholder in the process.

A procedure for stakeholder evaluation is suggested by Guba & Lincoln (1989:185), using a twelve step approach with multiple feedback procedures:

1. initiating a contract with the client or sponsor commissioning the evaluation;
2. organizing the evaluation;
3. identifying stakeholders;
4. developing within-stakeholder-group joint (collaborative, shared) constructions via the hermeneutic / dialectic circle process, specifically focusing on claims, concerns, and issues [CC&I];
5. testing and enlarging within-group constructions by introducing new or additional information and by enabling group members to achieve higher levels of sophistication in dealing with such information;
6. sorting out resolved CC&I – those on which consensus has been achieved;
7. prioritizing unsolved CC&I;
8. collecting information bearing on unresolved CC&I;
9. preparing an agenda for negotiation;
10. carrying out a negotiation;
11. reporting via the case study – the joint construction as product; and
12. recycling.

This approach is supposed to ensure continued stakeholder participation, as well as maximum benefit to the evaluators in ensuring that the evaluation grasps the most important issues at hand.

Some more points could be added to this list. One must be clear about the nature of the evaluation. If the true reason for evaluating is for the institution or program to improve their current practice, as organizational learning, then there are certain requirements to be met.

Feedback to practitioners is crucial (Weiss 1998). Through stakeholder participation, one has a better chance of reducing the feeling of control and increasing the level of acceptance throughout the organization (Knežević 2005). Similarly, "through an appreciation of *dialogue* in the pursuit of *reflection* and *learning*, the problem of evaluation […] can be addressed" (Knežević 2005:45).

The question of who is commissioning and who is conducting evaluation is not without importance. Today, most evaluations are commissioned from the top of the hierarchy, either from a level of government above the institution or from the top of the institution itself (Weiss 1998). For evaluation to be a true learning mechanism, it might be necessary to involve those being evaluated more.

> If participants can be taken seriously as observers, interpreters and judges of programs, why shouldn't their capacities to undertake their own evaluation studies be taken seriously? Instead of thinking in terms of evaluation as an activity orchestrated by evaluators, why shouldn't knowledge production about a program be linked explicitly to the collective objectification of experience and self-reflection which intimately informs and is informed by program practice? (McTaggart 1991:19)

The question of internal versus external evaluation is also important, although this is not clear-cut. With an eye to matters such as objectivity and autonomy, outsiders are preferable, but this must be balanced against concerns over the utilization of evaluation, intimate understanding of the present conditions and confidence among staff and administrators, all of which point towards using internal evaluators (Weiss 1998). Probably, a combination is best, especially with a certain stakeholder participation in mind.

2.3.3 Evaluator or researcher?

What role should the evaluator play? MacDonald makes a distinction between the evaluator and the researcher (MacDonald in Knežević 2005). Where "the *researcher* is free to select questions and seek answers to them […] the *evaluator* […] must never fall into the error of answering questions that no one but him or her is asking" (Knežević 2005:46). The researcher, out of professional interest or just plain curiosity, asks questions about the procedures of an institution or a program, and thus indirectly acts as an evaluator, while the

professional evaluator's job ideally is to broker information between the affected parts, to facilitate exchange and dialogue. This information may of course be used for research purposes.

3. Methodology

In this chapter, I present the main methodological issues surrounding the thesis. The chapter is divided into seven parts. The first part presents the research question, with some explanation as to why it is chosen. The second part presents the delimitations of the concepts used in my interview guides. The third part presents the general research design. The fourth part outlines the data collection method and the fifth the method of data analysis. The two last parts discuss issues of validity, reliability and ethical considerations.

3.1 Research question

My original research question as stated in my research proposal is "*how does the increased use of evaluation of educational programs and institutions in Norway and Finland affect the quality of these*"? As with most research projects, the research question has come to have a slightly different focus after the work has begun. I will here go through the original idea behind the research question, and how and why it has been altered since.

After working for some time with my fieldwork, I realized I had to make some changes to my original research question. In my research design, I ended up doing fieldwork in some Norwegian schools, interviewing key stakeholders about their experience with evaluation procedures, and doing interviews with some policy makers about their views on the effect of evaluations. To contrast the picture I got from these interviews, I gathered research on the situation in the Finnish education system.

In my original research question, I mentioned both educational programs as well as educational institutions. It quickly became clear to me that my fieldwork had very little to say about specific educational programs, and more to say about the case of several educational institutions in Norwegian municipalities, contrasted with the general policy in Finland. Any final research question would probably do well to exclude programs, and focus on the matter actually being covered.

I also thought I could say something about the effect on quality in education of evaluations. Now, this formulation seems a bit optimistic to me. The question of quality in education is vast and complicated, and it is difficult to state anything definite on it, especially when the main focus of this thesis is on the experience of stakeholders with evaluations in the school rather than the question of quality itself. The effect of evaluations, regardless of whether they are good or bad, might be very small on overall quality of the institution. Of course, the evaluation of the evaluations can give the researcher a clue as to whether the current policy is good or bad for quality, but it is difficult to give anything but recommendations for changing (or keeping) that part of the institutional operations. Thus, the final research question should include a more cautious phrasing with regards to quality, even though it will feature in the main body of the thesis.

With these reservations in mind, it is time to find a new formulation of my research question. It should state that I am looking into the use of evaluations in educational institutions, and the feelings towards these among stakeholders internal to the institutions. My revised research question looks like this: "*How is the procedure of evaluation conducted in Finnish and Norwegian schools, and how do the perceptions of staff and management within the organizations differ on the value and use of evaluation?*"

3.2 Delimitation of concepts

In my research, I have asked questions regarding specific concepts, such as evaluation, participation, stakeholder and suitability. As these concepts have been part of the common understanding between me and my interviewees, they need to be properly defined for them to be useful in the discussion of my findings.

By *evaluation* is meant an institutional process undertaken by a school where the school solicits answers to questions regarding the operation of the school from various stakeholders, such as parents, teachers and pupils. These answers are then collected, systematized and processed within the school, and the results discussed in some form.

By *participation* is meant formal inclusion in the process before, during and after an evaluation. This inclusion takes the form of either formulating questions, answering them, or being part of the process after the results have been processed, or any combination of these.

By *stakeholder* is meant anyone being affected by the evaluation. This means anyone taking part in the construction of, participation in or assessing of the evaluation or directly affected by the measures taken as a result of the evaluation.

By *suitability* is meant the construct validity of the evaluations, here taken to mean how well the questions in a questionnaire fit the local situation of the schools examined. The reason this word is chosen is for clarity for the interviewees. Although construct validity is more exact as concept, suitability is easier to relate to in an interview situation.

3.3 Research design

> A research design is the logic that links the data to be collected (and the conclusions to be drawn) to the initial questions of the study. Every empirical study has an implicit, if not explicit, research design. (Yin 2003:19)

Maxwell (1996) specifies five different parts of research design: goals, conceptual framework (here: theory), research questions, methods and validity. The idea is that the design should not dictate a linear structure on how to commence with the research project, since qualitative research is subject to permutations and changes throughout the project. This strikes me as a good way to set up a research design, albeit with some added elements. The research question and goal is outlined above, and the theory in another chapter, and here I present some more parts of the design of this project.

3.3.1 Sampling

In this thesis, I examine three Norwegian schools, all of them in the same municipality. The sampling process is semi-random. After doing a small pilot project in one of the schools, I contacted the municipal school management to see if some of the other schools in the municipality would be interesting for this kind of research. In order to make comparison

easier and to avoid problems of differing evaluation culture I asked for primary schools with roughly the same number of pupils in each. The municipal officials provided me with several schools to choose between that all fit the criteria, and I chose two more schools randomly from that selection.

The main advantage of this method is its ease of comparison. Because the schools were fairly similar in all aspects, I avoided many distorting factors that could influence the perceptions of the stakeholders involved. This includes the fact that all three schools were identical in the relationship between management and teaching staff: between the two was a small planning group. In the smallest schools, management is often part of the teaching staff, while the larger schools have a more formalized middle level.

However, the sample has some drawbacks. Firstly, sampling only from one municipality might mean that the specific political and administrative context of that municipality could have an undue influence on the answers. I hope to have countered this by including findings from other sources to complement my interview data.

3.3.2 Participants

Interviewed in this thesis are stakeholders in the schools: teachers, school management and pupils.

The *teachers* interviewed, 11 female and 4 male, had differing experiences as teachers. One had been a teacher for more than forty years; another had started that same year. They all taught in primary school, and were thus responsible for a multitude of subjects.

The three *principals* participating in the study were all male. They all had some form of school leadership training. One had been a principal for more than fifteen years; the others were fairly new to the job.

I also interviewed two *pupils* in one of the schools, one boy and one girl. They were both involved in the pupil's council, and were soon graduating to lower secondary school.

These participants were chosen because they were involved at the school level and at the school level alone. Although the thesis concerns itself with all levels of evaluation, the amount of empirical data must be limited to a manageable amount. It would have been preferable with data from the local and national bureaucracy as well as from the politicians responsible for making education policies in both countries. This could have provided a more complete picture, lending support to the discussion of the findings. However, I believe that the current data is sufficient to give a tentative answer to the research question.

3.3.3 Materials

Several materials were used in this thesis. Among them are evaluation reports from the Norwegian Directorate of Education, individual school reports, meeting minutes and previous research reports.

The evaluation reports from the Norwegian Directorate of Education were often just the collected and systematized answers from the questionnaires of the schools' evaluations. They present the answers of the schools in relation to both the municipality and the country average. Sometimes the reports come with recommendations for actions the schools can undertake.

The individual school reports consisted of the numbers from the school as gathered and analyzed by the schools themselves. The meeting minutes were from meetings in the schools' planning groups and plenary discussions of the schools' evaluation results. The research reports consisted of previous research done on the outcomes of evaluations in Norwegian primary schools, mainly Møller *et al.* (2006). All these reports and minutes are part of my findings chapter, along with the interview data.

3.3.4 Procedure

What is the best way to find an answer to my research question? How should I go about finding it? In order to come up with a satisfying research design to answer these questions, I

have to break my research question down to separate compartments. To answer the question, I must be able to do the following:

Give an indicator of how the use of evaluation is perceived on different levels in the educational system. The reasons for the formulation "perceived" is stated below. With this phrase, I put more weight on the opinions of those involved in evaluation than any quantitative measure of quality. Many have an opinion on the use of evaluations. To give an effective indicator, I have to identify the correct levels to place my inquiry: government, local bureaucracy, local institution management, individual teachers and parents.

To see what government thinks of evaluations, it should be sufficient to go to official policy papers. Even if a government is not following its own official policy, it is always interesting to examine the policy in light of actual praxis. However, it is at the institutional level itself that the actual evaluations are being conducted, and also where the effects are most pronounced. The main part of my research will take part at this level.

Show the differences and similarities between the two countries. In order to justify my research question, and to give a more complete picture of the situation, I also have to make a comparison between Norway and Finland regarding this issue. The research being done at school level in Norway is extensive, and duplicating the process in Finland would be too much for a thesis of this size. I must therefore rely on accounts by others in the situation in Finland. However, my main reason to include Finland is as a contrast to the Norwegian system. Finland has been chosen for two reasons: It is at the moment much talked about for its scores in several international tests of school quality, and it is also an example of evaluation routines that are different from Norway.

The aim is to produce a piece of work that answers the question posed, and highlights the effects of evaluation within an educational system, with a comparison with a different system to give an idea of what it could be like. This will not answer the question of quality in education, and not whether one system is preferable to the other, but it will provide at least some insights into the matter.

To do this, a cross-sectional research design was chosen. Such designs are usually defined by the following characteristics: They have more than one case, examined at a single point in time, with quantitative data to give patterns of association (Bryman 2004). This fits my research well, but with qualitative rather than quantitative data. More than one school is examined, at a single point in time[8] by interviewing key stakeholders with a semi-structured qualitative collection method.

3.4 Data collection procedure

For most of my data collection, I have chosen to use semi-structured interviews. The best way to find answers is to ask people (Kvale 1996:1). The semi-loose structure gives me a flexible yet clearly defined process for gathering data. Thanks to the structure, the interviews remain fairly similar, making it easier to pinpoint where the informants agree, and where they disagree. This is clearly an advantage when it comes to analysis, as it allows me to identify common themes and concepts. At the same time there is room for follow-up questions, clarification and a form of dialogue with my informants that is lacking in quantitative questionnaires. Establishing a dialogue also helps give the interviewee a sense of ownership in the process, giving my research a better chance of having effects beyond my own graduation.

For this thesis, I was interviewing professionals about certain aspects of their work situation. A central point of my thesis is the opinions of the people involved in evaluation activities, as this has a big impact on the result of the evaluation, and thus indirectly on institutional quality. I chose to take their words at face value, leaving out a lot of information on the interview situation: all the interesting little bits of information that help give a better impression of how it all went. All we have is the words of the informants, and these might of course be guarded, altered or outright untrue. However, it is just the words that we are after,

[8] Meaning I did not examine any before-and-after effects of reforms or changes, but rather looked at their modus operandi.

and in some way I feel this shows that I trust the informants to formulate things the way they like them to be.

My thesis is not just about the situation in Norwegian schools. As the research question states, I include a comparison with the Finnish educational system. For this part of the research, I rely on secondary data. This data will mostly be works by other researchers who have already analyzed this system. I also look at some official Finnish policy papers, to see what the official position on evaluation is. This allows me to paint a broad picture of this system to contrast with the more thorough description of the Norwegian educational system.

There are some problems with this method of data collection. Instead of getting the information first hand, I rely on the work of others. This leaves me to the mercy of their analysis and their omissions. Often, what is not said is just as important as that which is said. However, by supplementing the research (which of course will come from more than one source) with official policy statements I should be able to get a rather complete picture of the situation.

3.5 Analysis procedures

As my main analysis procedure I have chosen partial coding. This involves going through the words of the informants, looking for that which has informational value to the work, and seeing if I can find common themes that go through the different interviews. These common themes become the focal points when I tie the empirical findings together with the theory. The term coding is usually connected with grounded theory, but this is not such a project. Because of time and resource constraints, the coding has been done in fewer rounds, and without the same amount of concept saturation as is usually expected in grounded theory. Therefore, the method employed is one of partial coding.

Some things must be noted with this form of analysis. Partial coding means I am looking for things in the material that are pertinent to my research as I have defined it in advance. This means that I risk overlooking some nuances that in the first place don’t seem of relevance. By defining what I am looking for in advance, the process looks less exploratory and more

like a treasure hunt. I do not believe that this will impend on my analysis, though. Some projects are more open-ended, but for now I wish to identify some common themes within the data to hook onto the theoretical discussion.

Kvale (1996:189f) posits six steps in data analysis, the two last not always used, and not part of this thesis: The interviewee describes, the interviewee discovers new relationships during the course of the interview, the interviewer condenses and interprets meaning during the interview and sends it back to the interviewee, the transcribed interview is interpreted by the interviewer, a new interview may take place, and the interview might lead to some sort of action.

The first three steps happen during the interview itself. The interview is a situation of negotiation between the two parts over the meaning of the words being said. Here the interviewer is in the lucky situation to be able to take home the interview answers and analyze them in light of the other answers collected. The analysis is the fourth point, and is the only part of the process that takes up more time and space than the interview process itself.

3.6 Issues of validity

When constructing a research project, issues regarding the validity of the construct are among the most important that face the researcher. It is important that the questions asked are necessary to figure out the research question, while not including a lot of irrelevant information. In order to ensure validity, I have cross-checked my questions with other studies within the same field, most notably Møller *et al.* (2006), and found that my questions are largely overlapping with theirs in the fields that are touched upon by that study. Some of the other questions were collected during the course of a different research evaluation project undertaken at the Institute of Educational Research, and were constructed in tandem with two other persons. Similarly, some of the interviews were done together with another person. This should ensure that my personal limitations do not interfere in the construction of questions or conducting of interviews. Lastly, other forms of data collection than interviews, including evaluation reports, observations of meetings and meeting minutes,

ensures some triangulation of methods, a procedure necessary in order to overcome the limitations of a single data collection method.

Sapsford & Jupp argue that what has to be established is whether data:

> [...] *do* measure or characterize what the authors claim, and that the interpretations *do* follow from them. The structure of a piece of research determines the conclusions that can be drawn from it and, most importantly, the conclusions that *should not* be drawn from it (Sapsford & Jupp in Bell 2005:118).

Included in the appendix to this thesis is a sample interview guide. Although most interviews included many more questions than what was written down beforehand, it should be possible to judge whether the questions are relevant and pertinent to the research question.

This being a qualitative study with just over twenty interviews, there is no hope to make this thesis transferable to other circumstances. Any conclusion drawn from the findings from these schools is not automatically valid for other schools and other contexts. Thus, I do not consider the question of reliability in relation to this thesis. Hopefully, the general themes will be of interest to others.

A different validity issue deals with the ecological aspect of validity: how fruitful is it to compare across two different educational systems (Vogt 1993)? Certainly, there is a problem of comparison when one moves the analysis to compare between two different, closed systems. However, there are some arguments that make this permissible.

The first is that an international comparison is of great use when discussing a nationwide system like the Norwegian system of school evaluation. Only in this way can different approaches be found and weighed. There is also a question of whether there is a real, or just a perceived, difference between the two countries in terms of evaluation procedures. The last point is that the two countries are, in the big picture, not so different, operating within the same Scandinavian cultural sphere, with similar economic, ethnic and social characteristics. This is also one of the reasons why Finland was chosen as a comparative example for the Norwegian school system: they are so alike, and yet there are some notable differences.

3.7 Ethical considerations

Any research project involving persons is subject to some dilemmas of ethical character. The main issues deal with the relationship between the researcher and those taking part in the project and the questions of informed consent and confidentiality in interview processes.

3.7.1 Power in interview situations

There is always a question of the power relations in interview situations. The interviewer does potentially have a lot of power, as it is he or she who controls the information gathered. Not only does the researcher have all the data, whereas those interviewed only have access to the parts pertaining to their own interview, but it is the researcher who does the analysis of the data and puts it into a context. There is a danger that the final product is something different from what the participants had envisioned. A researcher must be sensitive to this, and maintain a measure of openness and honesty in the treatment of the data.

The asymmetric power situation between researcher and participant is never to be taken lightly. However, in the case of my research, I hope to have avoided the more serious problems related to this. All those interviewed, save the two children, where older than me, holding influential professional positions, earning more than me and having a longer education than me. This eliminates much of the power asymmetry in the situation.

3.7.2 Informed consent and confidentiality

According to the Norwegian National Committee for Research Ethics in Social Sciences and the Humanities (NESH), research projects where persons are involved should only commence “after the participants' informed and free consent” (NESH 2006:13). By informed and free consent is meant an agreement made without outside pressure or limitations to personal freedom, where the informant is duly informed of his or her part in the research project.

Similarly, any person taking part in research activity has the right to have any information about them or their personal relations kept confidential. This means that individual persons

should not be identifiable by the final research presentation, and the researcher should guarantee anonymity in the use and reuse of the data.

For this thesis, all interviews were given under written or oral consent and guarantees of anonymity and no record of their names are kept. The names of the schools and municipality where the research was done are also not mentioned in this thesis, and no information that makes it possible to identify them is included. All informants where informed that they were at any point in the process free to abort the interview or ask for their information to be taken out of the data after collection. In the case of the two pupils interviewed, special permission was retrieved from the parents before interviews were conducted. They were also informed of any possible consequences in taking part in the interviews and the process from interview to finished research work.

4. Findings

In this chapter, I will present the findings of my fieldwork in the Norwegian schools. I will go through the three main evaluations each of the three schools undertake on a yearly basis, and add the findings from a fourth evaluation that one of the schools has participated in this last year. For each evaluation, the main points regarding that evaluation will be presented. After this, I will report on the situation in Finland using current reports on evaluations there.

Before commencing, it is important to note certain qualifiers. A "finding" is not really what it might look like. I have not uncovered any secrets, or discovered anything that is a hard, incontrovertible fact. Rather, I have found that some issues are common for all the interviews I have made, and that this points to something that is really there. This something can be an issue that must be seen to, or something the institution has reason to be proud of.

It also means that nothing of this is new. Since all my information comes from people dealing with the school, be they teachers, principals or education bureaucrats, it is clear that there is already an awareness of these issues. This is therefore, as it should be, only a summing up and systematization of the information that is already within the institutions and system.

Also, the findings in this chapter are well in concurrence with other research on the perceptions of school staff and management on evaluations. Specifically, the School Leadership Survey of 2005 (Møller *et al.* 2006) shows that most of the observations herein fit with the current research on the use and perception of evaluations in the Norwegian school system. Research from Finland indicates that many of the same tendencies exist there (Kohonen 1996).

The chapter is broken down into sub chapters. As the results vary according to which evaluation is being examined, each evaluation will be discussed separately. However, some main themes run through all of them: the relation between management and staff, the use of evaluations, and the role of them in relation to school improvement. These general themes will be addressed under one heading. Along with the data from my interview sessions, I will present findings and analysis from the evaluation data itself.

4.1 The parental evaluation

The parental evaluation is mandatory for all Norwegian schools. It takes the form of an invitation to take part in an internet questionnaire distributed to the parents through the pupils. They are asked to log onto the web page of the Directorate of Education and answer some questions regarding communication with the school, the general satisfaction with the school and other issues concerning their role as parents.

For this evaluation, there are several findings to be discussed. The main points deal with the results of the evaluation, response rates, questionnaire relevance, post-evaluation process, and general use of the evaluation tools.

4.1.1 Evaluation results

Looking at the responses to this evaluation, the main result must be said to be that parents are satisfied with the school their child is attending. This is common for all the schools I have examined. Most respond positively to what the school is doing and no areas point themselves out as matters for worry. This corresponds well with the data we have from other inquiries into parental satisfaction at the primary level (Beck & Vestre 2007)[9]. Ironically, Norwegian parents tend to be more pleased with their local schools than the general level of satisfaction in the education system would imply (Nordahl 2000). This is an indication of how attitudes are shaped by both direct experience and more indirect means, like how media portrays the sector.

The main cause for concern is the contact between school and parents. The parents feel they don't always receive enough information on what the school is doing. However, they also feel that they have other means of contacting the school should they want to bring something

[9] An assessment of the same evaluation for a different school, with roughly the same conclusions, can be found here: http://www.oppvekst.nesodden.kommune.no/tangenasen/Meny/Foreldre/FAU%20filer/brukerus_rapport_2006.pdf, cited 28.04.08

up. While looking into the ways of communicating with parents might be a good idea, it is difficult to see this as a major problem the school should address immediately.

Of course, simply looking at the responses only answers half of the question. This evaluation has certain properties that makes it difficult to assess the value of it for the school, and hence the parents. Looking at response rates and potential methodological issues is just as important as looking at the findings themselves.

4.1.2 Response rates

This evaluation has a very low response rates. For all of the three schools I examined, less than half the parents answered the questionnaire the pupils brought home with them. In some classes only one or two parents responded. As only two or three class levels were taking the evaluation every year, this amounts to very few responses for each school for each year. The principal of one of the schools states this very clearly:

> We have had some problems with the evaluations this year. Only around 30 % of the parents answered the parental evaluation, and several of the classes didn't get to answer the pupils' evaluation because of technical difficulties.

There may be several reasons for this low response rate. The most probable one is that parents see little value in responding to the questionnaires. For one, they have other ways of communicating with the school. This can happen through the parental conference between parents and teachers, through the local Parent-Teacher Association, or by simple direct contact through telephone or letters addressed to the school.

It might also be that the schools do not inform the parents properly on the results of the evaluations after they have gathered information. This might lead to the level of trust in the value of the evaluation going down.

4.1.3 Questionnaire relevance

According to most of the teachers, the response rate meant that they did not put much stock in the results of this evaluation. One said that “very few parents answer the parental evaluation, so I haven't looked closely at that one. I don't feel it's very relevant when so few answer it.” This might mean that it is difficult to infer anything from these evaluations. There are several factors that could distort any conclusions based on these evaluations. With such a low reply rate, there is no hope of having representation in the answers.

Of course, the goal might not be to get a statistically representative evaluation, but simply to get some indication of what the parents think of the school. However, even with this modest goal in mind, the evaluation might not live up to expectations. As pupils must bring a note home from school informing the parents of the chance to participate in the evaluation, and the parents have to log onto the internet and answer the questionnaire, one should expect this to represent as a hindrance for answering, especially when there is no perceived gain from answering.

It is not improbable to expect a certain self-selection in the group of respondents. Only those parents who have strong views on some matter regarding the school is likely to answer, as they are already looking for an opportunity to give feedback to the school. Those parents with a satisfied, or even indifferent, attitude towards the school are less likely to respond. That is, unless the process is mandatory, involves some kind of reward, or is exceedingly simple to go through with. Some of the teachers mention that this evaluation could be taken in conjuncture with the parental conferences they have with the parents. As one puts it, “it might be possible to have questionnaires that the parents had to take at the conferences, to be sure that most would answer them.” The question of whether this is the way to go remains.

4.1.4 General use of the parental evaluations

The teachers express dissatisfaction with this evaluation, as it does not fit with their view of how such an evaluation should be done. This stems mostly from the low response rate, but also has to do with the teachers' perception of how contact with parents should be done. Most of them believe that direct, face-to-face contact with parents in the parental

conferences has a much greater effect, and is of much greater value, than the questionnaires. One puts it like this:

> We teachers see what is happening in the classroom, so it's easier for us to attend to any problems that might arise. It's very important with direct communication. I'm not a fan of forms and questionnaires.

Management is generally more inclined to use indirect measures of performance and satisfaction, but is still wary of using this evaluation as an integrated part of their development plans or post-evaluation work. Aside from the response rate issue, this probably has to do with the fact that most of the work the school does is independent of the parents. An institution like a school mostly takes its clues from the educational bureaucracy and political organs. Management might also feel that the use of the parental evaluation is limited. As one principal puts it, “the parental evaluation is taken seriously, but we are rather surprised that the parents listen more to common gossip than our information to them.”

Exactly what the “common gossip” is, I was not able to get an answer to. Nevertheless, this quote shows that the school in question feels it is distributing information to the parents that is not heeded, and that the school feels it is taking sufficient measures to bring the results of the evaluation back to the parents. Another principal says that in their upcoming development plan, “the focus will be parental conversations.” According to Møller *et al.* (2006), only about 26% of 615 Norwegian school leaders asked feel that the parental evaluations play an important role in improving the teaching methods at the school.

4.2 Pupils' evaluation

Every spring, the pupils at a school take part in an evaluation of several aspects of their everyday life at the school. This consists of an anonymous written questionnaire they fill out under the supervision of their teachers. It deals mainly with general well-being, problems with harassment, and questions regarding their workload.

The issues to be discussed with regards to the pupils' evaluation deal with the responses from the pupils, but also with certain methodological issues arising from the difficulties of using children's' opinions as a basis for action.

4.2.1 Evaluation results

As with the parental evaluation, this evaluation reveals that most of the respondents are satisfied with their schools. Very few pupils feel there is a problem with teasing and harassment, and the level of general well-being is high. This corresponds with other findings regarding pupil well-being in Norwegian schools. They like going to school.

Pupils also report that the amount of homework and difficulty of the school work is at an average level, with about half finding it fits well, and one quarter each find it too hard or too easy, respectively. That half of the pupils find that their work is too easy or too hard is an indication that the schools have not found the proper balance in their individual assessments. However, it also shows that the general level is neither too hard (where the answers would be skewed “downwards”) nor too easy (where answers would be skewed “upwards”).

4.2.2 Children as respondents

The pupils' evaluation is generally favorably assessed by the teachers and management. Most feel it is phrased in a careful way and with a satisfying degree of specification. However, several also express doubts as to the usefulness of the evaluations. This stems mostly from the problems concerning using children as respondents. Eder & Fingerson (2003) report on the specific problems of interviewing children. These problems are not alleviated by the use of questionnaires. If anything, the contrary is true.

This is expressed outright by several teachers and persons in management, with comments such as this, from a principal: “It is difficult with the pupils' evaluation, because pupils are preoccupied with the here and now, and this influences the answers they give.”

This is reflected in the general survey. Møller *et al.* (2006) find that only around 12% of the principals in a selection of 622 find the pupils' evaluation to be very influential in the development work of the school, with an additional 41% finding it moderately influential. These numbers represent an aggregate number from all three levels of Norwegian education. The survey authors note that this evaluation "has considerably lower support in primary schools than in lower and upper secondary" (Møller *et al.* 2006:85), but do not offer a break down of the numbers. This would suggest that the numbers cited are even lower for primary schools. However, primary schools are by far the largest group in the sample, with 72% of the survey schools being primary schools, so the real number is probably not that far from the ones presented.

Although this survey does not cover teachers' attitudes toward this evaluation, several of the teachers interviewed had reservations as to whether they could rely on the answers to be represented in a balanced way. One teacher said the following:

> Most of the pupils answer, probably because they sit and do it when at school. So I guess it can be called representative. But I don't know if it really matters for them. They might just answer anything to get it over with.

Even one pupil points this out: "The evaluations work OK, but not everyone is honest. I don't think that much comes out of it. The same people are always picked on." Again, there seems to be some uncertainty about how to deal with and interpret the results of evaluations. This could mean that the evaluation of how pupils experience life in school needs to be done in a different way.

I have not interviewed many children for this thesis. Mainly, this is because of concerns over these exact issues. However, before reaching this conclusion, I did conduct two interviews with pupils in one of the schools. These examples might themselves be good illustrations of the limits of the pupils' evaluation. The topic of the thesis concerns the use of evaluations as tools for improvement within an organization. For pupils not older than twelve, this is very abstract, and probably difficult to relate to in a meaningful way.

The answers I got were very vague, and the two pupils I did interview managed to contradict each other on several topics concerning the evaluations. One was happy with the phrasing of

the questionnaire, the other found it confusing. One felt there was not a problem with harassment, the other that several pupils got picked on regularly. One said that the evaluation did not feel anonymous, as they had to take it in the classroom with the teacher present, the other was very happy with the fact that no one could know what they had answered. And one said that the pupils were truthful, while the other said that they probably did not tell the truth.

Although this was only two pupils, and a larger sample might have nuanced the impression considerably, it is remarkable how divergent their answers were. Since I asked them less than ten questions each, that they would disagree on more than half of them surprised me.

4.3 Municipal workplace evaluation

The municipal workplace evaluation is taken each year, and is mandatory for all employees of the municipality. It deals with topics such as workplace well-being, relationship with management and general job satisfaction. It takes the form of an Internet form. It is the only evaluation examined in this thesis that is not provided by the Directorate of Education.

As this evaluation is concerned with mapping out workplace satisfaction for every single employee in a municipality, the whole thing takes a different form than the more school specific evaluations. However, this is the only formal feedback procedure the management of a school has as a tool to assess how teachers feel towards the school.

Issues discussed for this evaluation involve its concrete results, evaluation accuracy and phrasing, and the form of evaluation used.

4.3.1 Evaluation results

This evaluation is the most diverse, when looking at the results. For each of the schools I visited, different things were highlighted in the results. There are several reasons for this. Firstly, each of the three schools had different characteristics that influenced the answers. Two of the schools were only primary schools, while the third one included lower secondary

school. This means that the teacher composition was slightly different in that school. One of the other schools included an expanded after school program (SFO), meaning that there were more extracurricular activities employees there.

There was also a difference in how long management had functioned. In one of the schools, a new principal had just been employed, but the rest of management had been in place for a long time. In one of the others, the whole management was rather new, and in the last, the whole management had been working there for several years. This has an impact on the answers to this evaluation, as it deals a lot with employment issues.

In general, one can conclude that this evaluation showed that the employees are generally pleased with their work situation. There are no notable conflicts between the pedagogical staff and management, and what little dissatisfaction there is stems more from material concerns than interpersonal conflict. However, it must be noted that during the interviews several persons expressed concern over many issues they were unable to touch upon because of the rather non-specific nature of this evaluation. This general assessment is therefore based only on the questionnaire material it is possible to examine from each school.

In one of the schools, workplace satisfaction had increased notably in the course of the last couple of years. This reflects a long-standing conflict between pedagogical staff and the previous management at the school. For the other schools, the level of workplace satisfaction is more or less constant over the last years.

4.3.2 Evaluation accuracy and phrasing

The municipal workplace evaluation met with severe criticism from both teachers and management alike. Almost unanimously, it was criticized for being of little interest to the schools. Mostly this stems from the fact that the evaluation is a municipal one, rather than being designed for the school setting. The questions deal with the intra-organizational relationships in the workplace, focusing on the employees' perception of management, and do this for every workplace in the municipality.

The main problem is that all but the smallest municipalities have a large variety of positions, and every position requires a different approach. Therefore, any common evaluation of employee satisfaction faces a difficult trade-off between accuracy and inclusion. For every question relating to a specific practice in one kind of workplace, other parts of the municipality's activity is excluded. It seems that this evaluation has chosen to focus on including as many parts of the municipal workforce as possible. One teacher offers this explanation:

> The municipal workplace evaluation doesn't work at all. It's too general, so it's very difficult to answer the questions in any specific way. It concerns matters to do with leadership, including situations where it is difficult to specify what is meant. It might have some value if it shows something of clear significance, but the questionnaire is so generally worded that it's difficult to see what that might be.

The evaluation is supposed to fit every workplace in the municipality, and the responses of the teachers to both the phrasing and the set up of the questionnaire reflect this. Several teachers and management have reacted to the phrasings of the evaluation. Mostly, the reaction is that it is difficult to answer the questionnaire, as it is hard to relate the answer categories to the actual situation at the school. One of the principals has this to say:

> The municipal workplace evaluation is bad. The questions are too vague and too general. The evaluation has been going on for many years, but does not seem to result in anything concrete. Most things have been subject to evaluation, but there is seldom the right focus on them.

This evaluation might suffer from a lack of focus. It is meant to be a tool for the municipality to gather information on the workplace well-being of employees. Yet it is carried out entirely by the units themselves, and the results of the evaluation are returned to the units with recommendations from the municipal authorities. Managers of municipal units might face difficulties in handling this procedure. It is unclear what the municipality expects to get out of this evaluation.

In addition to the question of relevance, several teachers voiced concerns about the phrasing of some of the questions of the questionnaires. Relevance and question phrasing are tied closely together, as we see from the answers of the teachers. If the questions are not phrased

in a way that captures what the respondents wish to communicate, the evaluation relevance is damaged.

4.4 State of Affairs evaluation

One of the schools had recently gone through a fourth kind of evaluation, called the State of Affairs evaluation. In conjuncture with applying for project funding, the school was required to have an internal evaluation among the teachers of the school. For this, they were given time off from work to answer the questionnaire on-line. The evaluation was provided by the Directorate of Education, but bears little resemblance to other evaluations from the same governmental instance.

The questions in the questionnaire deal with four categories regarding learning outcomes, pupil well-being, and motivation and individually adapted learning plans. The answers were divided into three categories, traffic light color coded for reference, where green stood for "The school's practice is satisfactory and in keeping with the model picture", yellow for "The school's practice must be improved", and red for "The school's practice must be changed – measures are needed". Issues discussed in relation to this evaluation concern the results, phrasings and categories in the questionnaire.

4.4.1 Evaluation results

The main result from this evaluation is that the teachers are moderately satisfied with how the schools are working. Nine questions receive especially favorable answers, meaning where more than two thirds of the teachers answer in the green category. These concern pupil cooperation, the school's pupil council, meal time, sick leave routines for pupils, monitoring of drop-out tendencies, skills development at an early stage, teacher presence at recess time, parental involvement, and inclusion of other professional groups to solve pupil behavior related problems.

In twelve questions, less than a third of the teachers answer in the green category, but for only five of these is the red category of any size, where more than one in ten teachers

answers red. These five points are concerned with the use of digital tools in school work, pupils' ability to judge and use digital information, noticing and acting on slow pupil progress in subjects, pupils treating each other with respect, and general school maintenance. In addition, one question regarding teacher-pupil conferences received a red answer from one fourth of the teachers, but also close to half green answers.

4.4.2 Phrasings of the evaluation

This evaluation is worded in a different way from the other evaluations, both in the way the answers are graded, and in the way the questions are posed. This has been commented on by teachers and management alike, with opinions differing on whether this is good or bad. Mostly, though, the evaluation has been received favorably. This might have to do with the different procedure chosen for it, but might also reflect that the questions are better.

The principal of the school believes this to be the best evaluation:

> The State of Affairs evaluation is the best one. It touches directly on the everyday workings of the school. The others just barely touch upon everyday matter, but this one goes straight to the core. I like it a lot.

The main reason for this is probably that this evaluation has a narrower focus, with questions centering on very concrete issues in the school workday. This has been well received by all, as many teachers find the other evaluations to deal with matters too removed from real life at their school.

However, while the State of Affairs evaluation seems better suited for the school because of its focus, there are some caveats when it comes to the phrasing of the questions. As one teacher says,

> The State of Affairs evaluation was a bit unclear. We discussed the questions. It's important that teachers point out those questions that are so ambiguous that we or the parents don't understand how to interpret them. Not all the questions were good enough.

It is not difficult to see what is meant here. Of the 43 questions in the questionnaire, 10 contained conjunctions that made them double-barreled, and hence difficult to answer. Bryman (2004) states that avoiding questions that can be interpreted in more than one way is one of the most important tasks of anyone constructing questionnaires.

Several questions also contained words or phrases that are difficult to interpret because of their diffuse meaning. Examples include “making the values of the school come alive”, “active participation in society”, and “regularly”. These are difficult to relate a fixed meaning to, and therefore make the answers less clear. How often is regularly? How active is active?

4.4.3 Categories

In addition to having some phrasing issues, there is also a question of the use of the answering categories in this evaluation. The use of a color scheme is not normally used in these evaluations, but can function as an easy way to make the results more visible and understandable at first glance. However, the accompanying text is more confusing. Firstly, it is hard to see what the green category, “The school's practice is satisfactory and in keeping with the model picture”, actually means without any description of what the model picture is. Secondly, it is difficult to see the real difference between the yellow and red categories. One is “The school's practice must be improved”, the other “The school's practice must be changed – measures are needed”. Both mean that changes should be introduced.

The meaning of the categories is probably that there should be one for “satisfactory”, one for “working, but not perfectly”, and one for “unsatisfactory”. The way it is formulated, it is unclear how management should interpret the answers from the teachers. The problem is how to weigh the answers up. If there is much yellow, but hardly any green or red (as in question 13, with 9.1% green, 86.4% yellow and 4.5% red), how urgent is the change? If there is a moderate yellow and a moderate red (as in question 22, with 45,5% green, 31,8% yellow and 22,7% red), is this more urgent, or less?

One other interesting fact about this questionnaire is that, while this also is provided by the Directorate of Education, it is very different from the other evaluations from the same

directorate. This one is more local in focus, contains questions of a different nature, and has a different assessment system. There are no notable methodological problems with the other evaluations, while this one is ripe with them. One might ask whether this was designed inside the directorate or bought from someone else. Yet, this one was the most popular among both staff and management, so this might be the right direction to head for future evaluations.

4.5 General themes

Aside from the results and specific issues discussed above, the general themes that arise from all the evaluations are of equal interest. The comparison of the different evaluations reveals some similarities that should be addressed. These deal with more general observations concerning the workings of a school, but also with the strengths and weaknesses of this form of evaluation in general.

Most important among these general themes is the post-evaluation process, the roles of teaching staff and management in the school structure, the perception of best practice for improvement in the organization and the real and perceived value of this type of evaluation.

4.5.1 Post-evaluation process

> We need to bring in everyone, both teachers and management. Spend enough time on each evaluation and especially on the answers afterwards. We shouldn't do it in the most hectic periods of the year, because that will influence the results.

The above quote nicely sums up the importance of working with the evaluations after the results are returned to the school. Without a proper process after the evaluation, there is little hope of the school being able to use it as a tool for actual improvement.

The post-evaluation process is fairly similar in all schools. This is a result of how Norwegian schools are organized. The structure of most schools is pretty much the same for all schools, with some exceptions for the largest and smallest units. Generally, there are two main levels of employees in schools. There is management, normally consisting of a principal and some

persons working in assisting positions, such as inspectors and administrative staff, and teaching staff, with teachers, teaching assistants and extracurricular coordinators. Between these two levels there usually exists some kind of coordinating instance, most often a planning group. In larger schools, there might even be a more developed middle level, while in the smallest schools the management level is very small or even integrated in the general staff.

In all three schools examined, the process for assessing the evaluations after they are completed and the results are sent back to the school is the same. Management receives the results, and does a quick first-hand assessment of them. It is then taken up in plenary, where the results are reported, and discussed in detail in the planning group. If anything special is noticed and the planning group wishes to act on it, the evaluation results go back to the plenary, where teachers and management find a plan for continued action in workshops.

What happens after this point is more unclear. Most often, a document is produced highlighting the discussion points from the workshop, and this document is to be taken into account when the development plan for the year is designed. However, even though I have been given access to meeting motions and written reports, several of the teachers claim not to have seen any documents from evaluation processes. It is difficult for me to judge whether management is negligent in distributing information, or whether the teachers ignore or forget the information they receive and perceive as unimportant.

The main finding, mentioned by most of teachers I talked to, is that little happens after the evaluation results are processed. According to them, the process of evaluation has been going on for some years, but there is little to show for it. Generally, there is a feeling that the process of evaluation is mostly for show, and not for any substantial changes in the organization. The following three quotes from teachers in the different schools are in agreement:

> Not much has been done with the evaluations. Generally, they are not used widely after they are completed.

> The main use of evaluations is to get attention to problem areas, but they usually end as paper in a desk drawer. Nothing more happens.

> We go through the answers. Then we are split up into groups, who go through the evaluation and return answers to the principal. Then we don't hear anything more. I wish we could have something more concrete in writing. When the parents are involved it's always taken more seriously.

Of course, the picture is not this one-sided. Several teachers point to cases where things have happened as a result of evaluations. Examples include placing more teachers in the school yard during recess, and changing the way parental conferences are done. However, all the cases mentioned by the teachers interviewed came as a result of either a different evaluation not in use at the schools any more, the Olweus harassment prevention program, or from the parental evaluations. This gives some credit to the last claim in the third teacher quote above.

The post-evaluation process differs a little according to which evaluation is given. For the pupils' and parental evaluation, the process is the same: results come back to management, taken into plenary, discussed in the planning group and possibly taken back into plenary. For the municipal workplace evaluation, things are slightly different. Along with the results, the evaluating also comes back with ready recommendations from the municipality. These recommendations are usually of a very general character. These are taken up in the planning group, and might be included in the next development plan for the school.

The fact that this is overseen by the municipality instead of the Directorate of Education might have an effect. It is to the local authorities the results are reported, and they make recommendations to each workplace based on these. With the other evaluations, the schools are trusted to do their own post-evaluation work, finding out which issues are to be addressed. This might lead to more ownership to the process, and thus more commitment, but might also lead to less being done with the results.

In the school where the State of Affairs evaluation took place, several teachers commented that they were very pleased with the process both before and after the evaluation. All teachers were required to take time off from their work for an hour to answer the questionnaire, and did so simultaneously. After the results came back, the evaluation was subject to a week's project work in the school, with the staff holding its own post-evaluation

seminar and working with the results in groups. This was well received by the teachers, and several commented positively on this.

> With the State of Affairs evaluation everybody sat with their computers and did analysis together. It was very nice that we got the time to work with it in a proper way. Each got the results handed out [...] With the State of Affairs evaluation we sat in groups and all the groups were looking at the same thing at the same time.

One of the statements above has to do with the time dedicated to the post-evaluation work. Generally, teachers find that they have too little time to work on the evaluations after the results come in. This quote sums up the general sentiment:

> I miss being able to spend time on it, to talk about [what comes out of the evaluations], and do something about it. We never go in depth about it, there's always something new. [...] It usually ends up in the group discussion. There is some frustration over this.

This could either mean that the general workload of the school is too high, or that working on evaluations is not prioritized as much as other, more central tasks. After reviewing the answers from the teachers, it is difficult to determine which of these two explanations is correct. Most probably, it is a combination of the two.

When asked about the workload, several teachers respond that they have too much to do. Some lay the blame on just their management, others on the mandatory work required of the school, mandated from above. One teacher said this:

> There are a lot of projects and evaluations. Also, a lot of things come from high up, from the state and the municipality. [...] We have to keep moving things ahead in the calendar. This school year we were supposed to work with the Knowledge Promotion reform, and especially in the subjects Math and English. We still haven’t started on that, and now it's the end of the school year.

Several of the older teachers pointed out that working in school has always meant a lot of work. When it comes to the general workload of the school, many feel that it is a matter of how the school prioritizes it: “Management is pressed for time, so they will prioritize. They only do what they consider most important.” Management itself sees this as a case of priorities, as seen in the last sentence in this quote:

> Regarding time and resources, we feel we could use more of both. We have two full time positions in management divided among three people. This is sufficient, but we should have more time for pure office work. Now this accounts only for an 80 % position. For our main priorities we have enough.

4.5.2 The role of teachers and management

Central to the process of evaluation is the relationship between the different levels of the organization (Barr & Dreeben 1983). The assessment of things depends on the position of who is assessing. In schools, evaluations of the workplace are perceived differently by management and staff. This is seen in the schools I have examined. In the schools, there seems to be several disagreements over how things should be done or emphasized, both regarding the roles of each position and the assessment of evaluations.

As the schools are divided into three levels, the role of each level is subject to different interpretations. According to the teachers, their main devotion is to do a good job in the classroom, and the organization around this activity is simply there to facilitate this. This view is reversed when management is asked. According to them, the teachers constitute one part of a bigger organization, where everything needs to function in order for learning to be achieved. As one principal puts it, “the role of management is to look at the school generally, and see how the school functions internally. This way one can address issues with the relevant class levels.” This difference in views on important matters has consequences for how different aspects of the practice of evaluations are assessed by staff and management.

There is a certain disagreement between management and staff as to what should be emphasized in the evaluations. Some teachers feel that concern over parental dissatisfaction is having a too big impact on the school, while management tends to see this as natural. As seen above, one of the teachers states that “when the parents are involved, it's always taken more seriously.” This is confirmed by one of the principals, who says:

> Some teachers are annoyed that parents get involved. Management has more sympathy with the parents. The teachers tend to feel that parents are interfering with their work.

This shows that there is a discrepancy in the school's assessment of evaluations. Teachers tend to want to completely disregard the parental evaluations, while management is keen to be seen to take parents' opinions into consideration. This might have to do with the different opinions regarding teachers' and management's roles in the school. Teachers want to be seen as professionals, with a large degree of autonomy, while management wants to see the school as consisting of different stakeholders with an equal right to influence.

This difference in perspective also pertains to the evaluative process itself. One principal believes that while teachers are focused on the closed setting of the classroom, management is in a position to take the larger view:

> If the teachers are brought into the process they tend to deal with what concerns the children. Management deals with the things concerning the teachers. This creates a tension. In the State of Affairs evaluation the teachers focused on pupils and their conduct. We were given a totally different angle from the school director of the municipality. From him the message was to focus more on the teaching situation.

This is to some extent backed up by teachers who are part of the planning group. The following quote shows that position within a structure can have an impact on perspective:

> A major part of the problem is that we in the planning group know a lot, but the teachers don't. Especially when it comes to feedback. We feel we issue a lot of information, both by e-mail and orally. But in a busy workday the teachers feel they have to choose. The class is the most important, while the rest is more or less ignored.

This teacher is in a middle position in the institutional structure, in the planning group. Members of the planning group act as mediators between management and staff, and this teacher is here placed in a middle position regarding this issue as well. On the one hand, teachers are mostly focused on the everyday conditions of the classroom. On the other hand, this focus is a result of the teachers' workload and training, which forces teachers to prioritize within a strict schedule.

Another issue has to do with what the different groups find to be relevant to their situation. Several teachers mention the fact that they fail to see the connection between the evaluation work that is done and their workday experiences. One teacher in a planning group is sympathetic to this viewpoint:

> The development work isn't well known or close enough to the teachers. They concentrate more on their pupils. The close is more consuming. The development work done at the school isn't relevant to them in their workday.

A teacher not in the planning group confirms this, but is more condemning of the process:

> Another problem is the planning group. They have no authority. They're supposed to help the principal develop the school, not to act as labor union. Development depends on what the planning group is, and what commitments it has to the grade level teams. People don't show up at meetings and they don't have a relation to what the planning group is doing. They simply don't see the point. They only care about what goes on in the classroom. And they have seen time and time again that things are not followed up, that the work goes unfinished.

While it might seem that management sees the school as consisting of different, equally important parts, the issue is not this clear-cut. One principal says:

> We don't deal that much with the pupils. We understand that teaching is the most important factor in classroom interaction. This isn't well communicated to the teachers; we haven't succeeded in telling them directly.

This failure to communicate this to the teachers might have to do with how management perceives their role within the organization. While they agree that teachers are the most important part of a learning institution, they are often in opposing roles in relation to the teachers, and hence have to defend certain policies that the teachers do not like. This can lead to a situation where they do not communicate appreciation of teachers as much as they promote unpopular measures.

4.5.3 The perceived use of evaluations

Underlying all evaluation work is an idea evaluation is really a tool for improvement for the school. If it is not, then there is hardly any reason to evaluate in an environment where time and resources are scarce. However, evaluation serves other purposes as well. They can be legitimizing procedures for the schools involved, or a control measure from above, or even a combination of these.

It is difficult to find any unanimous agreement on the function of evaluations among the teachers I interviewed. Some of them pretty much dismiss the evaluations as paperwork for the bureaucracy, as seen in this quote from one of the teachers:

> Nothing new comes up, nothing we don't know about beforehand. If you don't know about the things that come up in these surveys, then you haven't done your job. The questionnaires [...] aren't very useful.

This particular teacher expresses a general distrust in evaluations, preferring to rely on the teacher's professionalism to inform the process of improvement. The reliance on personal ability to spot what is going on in the classroom is something shared among most of the teachers.

Others believe there is insight to be gained from participating in these activities. A teacher working in the same school as the one quoted above said the following regarding the parental evaluation: “The school can get an insight into how the school is doing in terms of well-being and learning”. This assessment was given even though this teacher generally feels that the parental evaluation is “not worth very much”. Another teacher stresses the information value of evaluations, even if they should not necessarily be used as basis for reform in the school: “The function is after all communication, it has the greatest value for those who evaluate the results of the evaluation”.

One interesting thing is that the general opinion of a given evaluation seems to reflect the opinions they had before they went into evaluation. This might be an indicator of how potential evaluations will be received, and also gives an idea of the motivational aspects of this work.

Among management, the question is more one-sided. All of the school leaders interviewed believed firmly in the idea of evaluations as an instrument for improvement, if not in the value of the specific evaluations themselves. The notion of evaluation as an *instrument* is interesting, as the school leaders interviewed tend to see the evaluations as just that. For them, it is a guiding paper, more than just a source of information:

> We identify problems, make them a part of our development plans, and then evaluate the result. This way we see the value of participating over some time in these evaluations.

The above statement can mean two things, depending on the perspective used. The first is that evaluations is used in an active way by school leaders, who carefully consider the results and act upon it, rechecking things as they make changes. The second view is that an evaluation is cause for more evaluations, and this is the source of a lot of problems.

It all narrows down to a question of what the real effect of evaluations is. It is beyond the scope of this thesis to establish the real uses of evaluation. However, it is possible to say something about how stakeholders perceive evaluations, and what they believe is the use of them. In the cases I have examined, it is hard to find a singular pattern in what teachers and management believe are the real uses of evaluation, as it is always tied together with other issues in the workplace, but the main perceptions have been described.

4.5.4 How best to improve

If evaluations are meant as a tool for improving the work done at the school, there is still the question of how this is best done. Improvement as a process has two parts: the first is identifying the better way of doing things, and the second is changing routines to make the improvement take place.

When it comes to the first question, there is some consensus among those interviewed that the evaluations are not necessarily the best way to identify best practice. The most striking feature is that teachers and management alike believe direct contact with pupils and parents to be the best way to both gather and distribute information. A teacher says this:

> Conversations with the pupils must be the basis for such things. It's much easier to know whether things are fine when you speak face to face. The same goes for parental talks.

This is a good summing up of the general sentiment. “Face-to-face” seems to be the catch phrase: “Contact with the parents is best maintained in face-to-face talks, at the parent-teacher conferences”.

This insistence on personal contact has its flip side in a distrust of information of a more quantitative sort. Several teachers mention that they do not trust questionnaires, and that they should be their own source of best practices: “I'm not a fan of forms and questionnaires. A teacher should have knowledge of the situation in the classroom, or he or she is not doing their job”, as one put it. However, many mention the need for some kind of anonymous feedback procedure for both parents and pupils, as it might be difficult to address personal issues in direct communication.

Similarly, the principals agreed with the teachers in that the single most important factor in identifying problems and solving them is the interaction between the teacher and the pupils: “It's very important with direct communication”. Where the two groups disagree is on the added value of evaluation forms.

The second part of improving practice is in the actual application of the information gathered. Again, there seems to be a disagreement over whether this is done properly. Mostly, management tends to think that when the overall schedule and number of tasks of the school is factored in, enough time is set aside for evaluations. Conversely, many teachers complain about lack of time, and wish more thorough work could be done with the evaluations:

> We should have a very own theme period of evaluation. A project where we set aside enough time before and after the evaluation and a deadline for when the results should come. We should also set goals for what is to come of the evaluation. People aren't informed ahead. You shouldn't feel like you get stuff thrown after you, because then people simply boycott it. They aren't given ownership of the process.

4.6 Finland

In order to get a broader view of the situation anywhere, it is always useful to look at what is going on in other countries. The comparative perspective opens up for an analysis between differing systems, allowing for a discussion of strengths and weaknesses, as well as an idea of whether the characteristics of a system are endemic or more universal.

In this sub-chapter, I present a literature review on the state of evaluation in the Finnish educational system. As I have not conducted my own first-hand research in Finnish schools, due to the limits of the research project relating to time, space and language issues, this part will focus on second-hand research from other sources. I still believe the end result will be satisfactory, and suffices to provide valuable insights and a useful counterpoint to the discussion of the Norwegian school situation.

The findings from Finland are not unanimous in any way, but it is still possible to point to certain features that both separate it from Norway and show the basic similarities between the two systems. The main findings are presented here.

4.6.1 Teacher professionalism and trust

Evaluation can not be seen as separate from the role of those employed in schools. How teachers and management see themselves will affect their attitude towards evaluation. Thus, how teachers see their professional situation will have an impact on the reception, execution and assessment of the evaluation and the evaluation process.

The Finnish educational system produces teachers with a high degree of professionalism (Webb *et al.* 2004). Not only are Finnish primary school teachers required to have at least a master's degree from a university, they are also given a large amount of individual responsibility for their teaching. One of the things that have helped advance this professionalism is belief in having a research-generated knowledge base (Westbury *et al.* 2005). That is why all Finnish teachers are required to write a research thesis as part of their training.

This professionalism has the effect of increasing the independence of the whole school. For example, it allows for more room in evaluation situations: “Each school, at least in theory, could design its own change strategy with mission statements, vision and implementation methodologies, and schedules” (Sahlberg 2007:157).

Additionally, the Finnish school system enjoys a high level of trust from the public:

> The culture of trust simply means that education authorities and political leaders believe that teachers, together with principals, parents and their communities, know how to provide the best possible education for their children and youth. In Finland, this transition from bureaucratic central administration to a decentralized culture of trust happened during deep economic crisis and public budget cuts in the 1990s (Aho *et al.*, 2006). It was argued that this happened because state authorities did not want to make difficult financial decisions that would cut local education budgets and thus have negative effects on schools (Sahlberg 2007:157).

As stated, the decentralization of the Finnish educational system could be the result of politicians afraid of making unpopular decisions. In this case, it seems to have had some good side effects.

4.6.2 Accountability and control

There is little use of external evaluation of any sort in Finland. For pupils, the only use of evaluation is the matriculation exam, under the control of the national examination committee appointed by the Ministry of Education, and tests taken among sixth and ninth graders, administered by the National Board of Education.

The point of these evaluations is to give teacher and pupil an idea of the pupil's standing in the subjects tested, and provide the pupil with a motivation to improve. However, some researcher indicates that these evaluations are

> being used to serve a multiplicity of purposes for which it was not really appropriate, of which an increasingly important one was accountability to the municipalities. However, there was concern that [evaluation] information might be used for comparing schools and teachers, resulting in competition between schools and teaching to the tests. (Webb *et al.* 2004:98)

If evaluation information is used for competitive ends rather than improvement, this could challenge what is described as an equity framework for accountability, where evaluation is used as a tool to redirect resources to where it is needed most, with the aim of equity rather than equal opportunity (Itkonen & Jahnukainen 2007).

In the literature on Finland, the twin concepts of self-evaluation and internal accountability are often mentioned. They both stand to mean a form of evaluation where the institution itself has at least partial responsibility for evaluating practice. Ideally, this should lead to

smarter and more efficient evaluation within schools, leaving to the political authorities the responsibility of facilitating evaluations and subsequent changes in policy and practice. With the national curriculum reform of 1994, Finland decided to decentralize responsibility for developing curricula, and also for evaluation.

According to Sahlberg,

> This has created a practice of reciprocal, intelligent accountability in education system management where schools are increasingly accountable for learning outcomes and education authorities are held accountable to schools for making expected outcomes possible. Intelligent accountability in the Finnish education context preserves and enhances trust among teachers, students, school leaders and education authorities in the accountability processes and involves them in the process, offering them a strong sense of professional responsibility and initiative (Sahlberg 2007:155).

Apparently, this has led to very positive results for Finnish schools. However, this culture of self-evaluation has met with mixed response from education stakeholders.

It turns out, as is predicted by the stakeholder approach, that schools respond favorably to having a say in the assessment of their own work: "schools had ownership over their methods of data collection and analysis and commitment to respond to evaluation findings which led to direct and immediate changes to practice" (Webb *et al.* 1998:554). This is totally in accordance with the ideology of the stakeholder approach.

Yet, all is not working flawlessly. According to Webb *et al.*, one of the major problems with the self-evaluation approach is that they

> appeared to have considerable limitations—for example, the data collected tended to be on a wide variety of often unrelated aspects of schooling gathered through various techniques and with wide variations in rigor. [...] Schools were viewed as hindered in their approaches to self-evaluation by lack of evaluation knowledge and insufficient financial resources (Webb *et al.* 2004:99).

It seems like schools might have a problem properly delimiting the subjects of evaluation, and that they tend to rely heavily on questionnaire data, with all the connecting problems. While Finnish teachers are teacher professionals, this does not mean that they necessarily are

evaluation professionals. External evaluators might have a part to play in the evaluation process, even if one is pursuing a stakeholder approach to evaluation.

Webb *et al.*'s study of evaluation in Finnish schools and its follow-up study six years later noted a change in attitude towards self-evaluation:

> Our follow-up study found that the initial enthusiasm for school self-evaluation in Finland had dissipated and was being replaced with skepticism as to its usefulness. For example, the teachers at Makilampi were generally supportive of the notion of school self-evaluation and [...] were considering introducing aspects of peer evaluation to allow the sharing of strengths and weaknesses in their classroom practice (Webb *et al.*, 1998). However, six years later the head teacher considered that the process of self-evaluation had become a superficial exercise. (Webb *et al.* 2004:100)

This obviously presents a challenge to the stakeholder approach to evaluation, a challenge that can not be taken lightly.

4.6.3 Decentralized educational system

A movement towards decentralization of the educational system can be observed in tandem with the move towards increased teacher professionalism. As seen above, this might have to do with politicians trying to wash their hands of the responsibility for whatever goes on at the local level, but it has had the effect of creating an autonomous and highly heterogeneous education sector in Finland (Westbury *et al.* 2005).

This again feeds back into the autonomy and professionalism of teachers, as they now have responsibility for not only curriculum, but evaluation as well. This sort of participation is welcomed by school stakeholders:

> Such decentralization has been seen positively by many teachers: it foreshadows increased participation in, for example, managerial and curricular decision-making, thus opening new options for shaping a distinct school-specific policy (Westbury *et al.* 2005:476).

However, the same issues arise when discussing decentralization as when discussing self-evaluation. It is unclear whether the schools manage their increased responsibility in any

constructive way, or if they end up wishing for someone to come and give them clear directions.

4.6.4 Lack of New Public Management

Many have noted the fact that although Finland went through a certain degree of deregulation and market exposure through New Public Management, like other Western European countries during the 80s and 90s, the education sector was left largely alone:

> although the emergence of the new public sector management meant revolutionary changes in Finnish educational discourse, this new rhetoric and practices have not been able to take root in education as easily as in other parts of society (Sahlberg 2007:157).

Also, the process of New Public Management reform is more limited in Finland than many other countries. It is notable that those public reform processes that have taken place have been more oriented to suit the public sector rather than privatizing previously public tasks:

> The Finnish public management reforms of the decade from 1987 owed most to the thinking of senior public servants. External participation from business people or consultants was the exception rather than the rule (Pollitt & Bouckaert 2004:58).

Although there has so far been less of those features that make up typical New Public Management reforms, "signs are growing that the Finnish education system will soon be expected to devote increasing attention to efficiency and productivity" (Sahlberg 2007:168).

5. Discussion

In this chapter, I present some reflections over my findings and couple these with the theoretical perspectives described earlier. The discussion will center on five points, each with its own sub-chapter. The first deals with the use of evaluations as legitimizing measures, a procedure undertaken more to show that something is being done rather than aid in improving the schools in question. The second part discusses the findings from Norway and Finland in light of the theoretical discussion of bureaucracy and control. The third part deals with the ideal evaluation as seen and negotiated within schools. The fourth deals directly with the problems of different roles within an organization, and the ensuing power conflicts. The last part deals with the problem of stakeholder evaluation versus evaluation expertise.

5.1 The form of evaluation

As seen in the findings chapter, both Norwegian and Finnish teachers find some of the aspects of evaluations problematic. There are some problems concerning the type of evaluation most often chosen in the two countries. One of the most important problems is related to the *form* of evaluation. There are several ways to conduct an evaluation, but a main division between them is the difference between external and internal evaluation.

Norway has external evaluation in the sense that the evaluations are designed by an external agency, most commonly the Directorate of Education, and commissioned by a different external agency, the Ministry of Education. Even though they are executed at the local level, all results are reported back to the external organ. This process is the same for all schools. It shares many traits with the classical approach to evaluation.

This form of evaluation has both positive and negative effects. Among the positive is that it is of higher use to the surrounding bureaucratic system because the information is gathered directly from the school data without first going through the internal interpretation process. It also facilitates comparison between schools because the data is exactly the same for all

schools, regardless of geographic location, school size and municipal rules and regulations. This ease of comparison also benefits those looking for trends over time in the questions asked in the questionnaires of the evaluations.

More negative is the loss of institutional autonomy and participation, as well as the feeling among many who work in schools that this is an undue interference in their professional business. In Norwegian schools, teachers are unhappy with what they see as a chore, taking time away from their core task of teaching. With the evaluations coming as a mandatory part of school operations, teachers and management feel there is little room to do things the way they would like to do them. External evaluation leads to little feeling of ownership over the process, and may lead to reporting and follow-up of results not being prioritized enough. This again lowers the usefulness of the evaluations. Additionally, there is little room to customize the evaluations so that they fit the individual school. This means that although the questions asked are the same, and the data therefore should be comparable, the different context of schools means that the validity of such a task is compromised.

Finland has a stronger tradition of local evaluation, with the schools responsible for choosing their own form of evaluation. The Finnish education bureaucracy places few restraints on the type of evaluation chosen by schools, and reporting to evaluation authorities is not mandatory. This model shares many traits with the stakeholder approach to evaluation, putting much store in participation and ownership.

Regarding Finnish schools, we saw in the findings chapter that both staff and management in Finnish schools welcome the freedom the decentralized model of evaluation offers, seeing it as an opportunity to reflect over the practice of the school in a way that all parts of the institution can take part in. Similarly, this alleviates the bureaucracy of some work, leaving it to focus on providing the schools with the assistance they might need to achieve the goals that are set.

However, this model also entails some problems. Mostly, these problems are the opposite of those associated with the classical, external evaluations. Firstly, the decentralized model of evaluation makes for a highly heterogeneous, often chaotic body of evaluation data. This means that comparison between schools and identifying the strengths of one school's

approach is increasingly difficult. Also, research into the matter show that Finnish schools are not necessarily succeeding in making any real changes to practice as a result of this approach to evaluation, as staff and management might not have the required knowledge or expertise to conduct a well designed stakeholder evaluation.

5.2 Bureaucracy, control and intra-organizational roles

The different regimes of evaluation feed into the discussion of bureaucracy and control in the theory chapter. The question is always to find the balance between bureaucracy and organizational freedom. While the bureaucracy is there for a reason, and can often have a positive function in systematizing information and procedures, one must be careful to avoid its tendency to want to make everything a case of reporting upwards in the system.

Stakeholders in Norwegian schools often feel they are being forced to prioritize bureaucratic processes over their core functions. In this environment of high workloads and low work satisfaction, evaluation can come to be seen as yet another useless task laid upon them. This created a negative association with evaluations, making it even more difficult to use them in a positive way.

Finnish schools have fewer mandatory tasks given to them by the bureaucracy, and the teachers have an identity as independent professionals. This has the positive effect of freeing up work for the bureaucracy, who can focus on other improvement measures, or even be slimmed down. However, many feel that the evaluations are not leading to much, and there is a feeling of disenchantment towards the current decentralized system. This in turn makes the arguments for more centralization seem more compelling.

All this is also a question of how a school should function. Although there can be no doubt that the core function of a school is giving children the tools they need to function as human beings in society, there is considerable disagreement on the roles of staff and management within schools.

This is supported by statements from both teachers and principals. There is general consensus that the core function of a school lies in teachers in the classroom. The difference lies in how teachers can do their work optimally. Teachers in Norway and Finland alike value the ideal of independent, professional teachers. The difference between the two countries lies in the actual amount of control required by the bureaucracy.

Norwegian school leaders seem to both want much control and be controlled from above. The number of mandatory evaluations for each school is rather high, with three evaluations of school practice every year in addition to national and regional tests of pupils' achievements. Also, school management is reluctant to let teachers influence the evaluation process, sometimes even skipping plenary treatment of evaluation results. Norwegian school leaders have considerably more trust in the chain of command as regards evaluation than their teachers.

Somewhat differently, it seems like Finnish school leaders have been giving their teachers considerable independence in setting up teaching and evaluation. Although the satisfaction with this system is still high, there is a movement towards more standardized evaluation, in order to give stakeholders more guidance in the evaluation practice and produce more robust data to use in the national evaluation work.

There is always tension between the managing and the executing levels of an organization. This shows in the context of evaluation too. Clearly, evaluation can be used as a weapon in the struggle to control the direction of an organization. Right now, there is considerable dissatisfaction with the current model of evaluation in Norway. This might signify that a change towards more stakeholder participation is coming. In the other end of the scale, Finnish school managers and bureaucrats are vying for more control of schools. We might see a movement towards convergence of the two systems in the future.

5.3 Face-to-face interaction as communicative action

Within both education systems the practitioners put a high value on personal contact with stakeholders. This goes for teachers and management alike. Presumably, and in accordance

with the Habermasian notion of communicative action as a form of negotiating towards a common understanding of necessary action, this is a more valuable form of evaluation input than any questionnaire results. Basically, humans are social beings, taking our cues more from interaction with others than from numerical data.

This form of communicative action is valued in all levels of the evaluation process. Increasing communication between teaching staff and management might alleviate some of the tension that naturally builds up between these two levels, and might lead to a higher degree of ownership to the process. Similarly, increased communication between other pupils and parents and the school represents a direct and unmitigated form of information, a clear advantage over the information obtained through the formal evaluation process. This means that, unless the information is factually wrong or someone is directly lying in their dealings with other stakeholders, direct communication is significantly more efficient than questionnaires.

The last point can not be overstated. The formal process of evaluation is costly and time-consuming, while a simple conversation with teachers or parents is quick, cheap and can potentially give a lot more information. Developing the best possible ways of communicating with both internal and external stakeholders will most likely produce results fast and cost efficiently for school managers and education officials.

5.4 The need for external guidance

As seen in the findings chapter, both Norwegian and Finnish education professionals have limited knowledge on how to design, conduct and assess evaluations. This has implications for how evaluations can be done, as the need for evaluation professionals is clear.

From Norway we see that while the evaluations in most cases are designed by someone with the required skills, the schools lack the ability to use the evaluation results properly. The result is that many evaluations are not used optimally, and in some cases not used at all. This way they serve as purely legitimating procedures without any real value as instruments for

improvement. If schools were able to assess the evaluation results in a proper way, some problems might be taken care of in a more timely fashion.

Similarly, although Finnish schools have different ways of conducting evaluations, the professional staff and management are not necessarily in a position to utilize their freedom to design evaluations. Even if they are at liberty to decide what kind of evaluation they want to have, and how it should be done, they tend to use it for largely irrelevant purposes (Webb *et al.* 2004).

The question is whether it justifies a move towards more centralized evaluation (for Finland) or more democratic participation (for Norway), and whether this requires more external evaluators or more evaluation training for school professionals. As there are few evaluation professionals working in schools in both Norway and Finland, the short term solution is probably to start working more closely with external evaluators when conducting evaluations in schools.

One seemingly successful approach utilized the insights of Guba & Lincoln's so-called Fourth Generation Evaluation (FGE) to empower internal stakeholders to do evaluation while being guided by external evaluation professionals (Lay & Papadopoulos 2007). They embrace the use of "constructivist" methods in otherwise rigid procedures of evaluation:

> Following constructivist principles provided an excellent opportunity for the different stakeholder groups to reflect on the project and deepen their knowledge and understanding of one another's perspectives and values regarding it. It also provided a framework through which they could jointly and collaboratively contribute to the project's ongoing development. [...] Projects and programs under highly autocratic, hierarchical management systems could benefit most from the empowering effect of FGE. [..] Our own experience supports this view; however we also recognize that attempts to conduct a constructivist evaluation are likely to fail unless the preconditions for it prevail (Lay & Papadopoulos 2007:503-504).

The preconditions necessary for this approach to prevail is most notably the use of skilled evaluators coming in to guide the process:

> [...] it requires very careful and skilled facilitation by the evaluator [...] in order to facilitate consensus between stakeholders with often opposing agendas. As Guba and Lincoln (1989) point out, finding evaluators skilled enough to undertake a FGE is not always easy and training may be required (Lay & Papadopoulos 2007:503).

This approach clearly takes some of its cues from the concept of communicative action and its evaluation offspring, the stakeholder approach. Yet it stresses the importance of outside guidance in the design of evaluations, which, as we have seen when examining school stakeholders' ability to utilize evaluative freedom, is of the highest importance.

Over time, it seems, it might be a good idea to start training some of the teachers or administrative staff in the design and execution of evaluations. This might facilitate better information exchange between the different levels of the education bureaucracy, and possibly lead to better results over time.

5.5 Evaluation and New Public Management

The classical approach to evaluation is decidedly in line with the neoliberal ethos of the system of New Public Management, as it is based on positivist epistemology and liberalist individualism. This is something both proponents (Reynolds & Teddlie 2001) and critics (House 1981) can agree on. The question is whether the current international trend of centralizing bureaucracy and decreasing teacher autonomy makes stakeholder evaluation reforms possible.

Even where the introduction of more stakeholder sensitive evaluation could be wanted, like in Norway, the general trend shows that this could be difficult to achieve. Not only is Norway part of the “second wave” of evaluation, where evaluation is ever more institutionalized (Boyle and Lemaire 1999), but implementing a change in evaluation towards more stakeholder involvement would require changes throughout the educational bureaucracy.

Firstly, while the Directorate of Education would still play an important role in training evaluators and teachers in evaluation, they would have to give up control over the

construction of evaluations. As stakeholder evaluation calls for more participant autonomy, both construction of evaluations and interpretation of the results are placed in the hands of those participating. This means less uniform data for higher level comparisons, one of the main goals of the classical evaluation movement (Reynolds & Teddlie 2001).

Secondly, this opens up for more negotiation of content and form of evaluations in the schools themselves. Today, this process is dictated more from above, giving school management power over the process. One of the main features of New Public Management is its focus on recentralizing governance. Thus, introducing stakeholder evaluation goes against the very basis of this system of management. Its prevalence means that any movement towards stakeholder evaluation might be unlikely.

In Finland, we see a move away from the evaluational autonomy practiced since the end of the 1980s (Webb *et al.* 2004). Although the level of institutional autonomy has been lauded, there are forceful arguments to include some form of professional guidance in the process (Sahlberg 2007). This fits well with the general trend of New Public Management, and there are signs that this governance system is on the rise in Finland (Pollitt & Bouckaert 2004). The first of these signs is seen in the strengthening of public sector accountability the last ten years (Holkeri & Summa 1996).

The New Public Management form of governance has come to mean centralization of bureaucratic procedures and a reinforcement of management over autonomy. This system of administration is prevalent in all modern welfare states, to a varying degree. In the case of Norway, this can stand in the way of a necessary change in evaluation practice. For Finland, the possible need for increased use of external guidance in the evaluation process coincides with the rise of New Public Management, a process that may augment the move towards more classical evaluation, regardless of whether the introduction of the administration system is beneficial or not.

6. Conclusion

In this chapter, I try to sum up the points from the discussion, adding some thoughts on what this might imply, both theoretically and in practice, for the two countries examined, and for evaluation practice in general.

6.1 Main findings

Although there are many minor findings in this thesis, especially concerning the fieldwork done in Norwegian schools, there are a few points that stand out as more interesting in a comparative and general setting. They mainly deal with the ways to conduct evaluation and the effect of different organizational and bureaucratic cultures.

This thesis has showed that there are two approaches to evaluation, the classical approach and the stakeholder approach, both with their strengths and weaknesses in practical life. These two approaches differ in many ways: their views on what and who evaluation is for, how it should be done and who should be a part.

It also shows that there are different approaches to evaluation depending on ones position in the system of evaluation. Teachers view evaluations differently from managers, who again see them differently from bureaucrats. Similarly, a difference in organizational culture matters as to the reception and utilization of evaluation results. Where the system encourages organizational independence, the negotiations of evaluation are less on the value of evaluation itself as in the use of results, while in a more closed system the process of evaluation is itself the focus of much conflict.

6.2 Implications

The findings of this thesis have some practical and theoretical implications. The discussion on how evaluation is being performed, negotiated and assessed points to two things: firstly, it

gives rise to a view that evaluation is not a static thing, always conducted in the same way with the same methods, but rather a continuum of methods and choices available to evaluators. The trick is to find the correct balance between control and comparability. Secondly, it is possible to start the discussion on what an optimal evaluation practice might look like in any given situation. This is, of course, far outside the scope of this thesis, but it is nevertheless interesting to look at some of the relevant points.

6.2.1 The balance of evaluation

There is a difficult balance to achieve between the increased, and often necessary, checks and balances, control and (at least superficial) accountability offered by classical evaluation, and the increased democratic legitimacy, sense of empowerment and often increased construct validity offered by the stakeholder approach to evaluation.

Instead of seeing evaluation as two distinctive approaches with strictly defined limits, it might be useful to see it as a continuum, with the strict, positivist classical approach in one extreme and the flat-structured, interpretative stakeholder approach in the other. Moving along this continuum, for example from the classical end towards the stakeholder end, means gaining some ground in terms of democratic legitimacy, perspective plurality and participation while at the same time losing some comparability and accountability. The perfect balance is most likely impossible to achieve, and will differ according to context, but starting to think in terms of this continuum might make the task of choosing the right evaluation design easier.

6.2.2 The ideal evaluation

After a discussion of the main approaches to evaluation, it is tempting to envision an ideal of evaluation. It is also tempting to adopt a middle-of-the-road strategy, combining the detached classical evaluation with the stakeholder approach. How might such an evaluation look?

With the stakeholder approach in mind, an evaluation might in a perfect world consist of one or several internal evaluators with good knowledge of the program or institution in question to include all involved parties as decision-making partners and one or several external evaluators acting as objective alibi and with a more research oriented approach. Together, they could equalize each other's weaknesses and amplify their strengths.

Although there might not be such a thing as the "ideal evaluation", since it is such a heterogeneous and diverse field and so dependent on context, the example of Lay & Papadopoulos show that a successful combination of classical, external evaluation and the stakeholder approach can work to improve how evaluation is done. This in turn can improve how evaluations are used to change practice, creating a chain of improvement, which ultimately leads to an improvement in practice and a positive feedback process for further well-balanced evaluation. That is, in an ideal world, at least...

References

Aldrich, Howard E. (1992) “Incommensurable paradigms? Vital Signs from Three Perspectives”, in Reed, Michael & Hughes, Michael (Eds.) *Rethinking Organization: New Directions in Organization Theory and Analysis*. London: Sage Publications.

Barr, Rebecca & Dreeben, Robert (1983) *How Schools Work*. Chicago: University of Chicago Press.

Beck, C. W. & Vestre, S. E. (2007) *Foreldres oppfatning av grunnskolen* [Parents' view of primary education].

Bell, Judith (2005) *Doing Your Research Project: A guide for first-time researchers in education, health and social science.* McGraw-Hill International.

Boyle, Richard & Lemaire, Donald (Eds.) (1999) *Building Effective Evaluation Capacity: Lessons from Practice*. Transaction Publishers.

Brand, A. (1990) *The Force of Reason*. Sydney: Allen & Unwin.

Bryman, Alan (2004) *Social Research Methods*. New York: Oxford University Press.

Carnoy, Martin (1999) *Globalization and educational reform: what planners need to know*. Paris: International Institute for Educational Planning.

DiMaggio, Paul T. & Powell, Walter W (1983) “The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields”, in Nygaard, Claus (Ed.) *A Reader on Strategizing*

Duncombe, William; Miner, Jerry & Ruggiero, John (1997) “Empirical evaluation of bureaucratic models of inefficiency”, in *Public Choice*, Vol. 93, pp. 1-18. Kluwer Academic Publishers.

Eder, Donna & Fingerson, Laura (2003) “Interviewing Children and Adolescents”, in Holstein, James & Gubrium, Jaber F. (Eds.) *Inside Interviewing: New Lenses, New Concerns*. Sage Publications

Foucault, Michel (1980) *Power/Knowledge: Selected Interviews and other Writings 1972-1977*. Brighton: Harvester Publishing.

Fraser, Nancy (2005) “Transnationalizing the public sphere”, http://www.republicart.net/disc/publicum/fraser01_en.htm. Vienna: European Institute for Progressive Cultural Policies

du Gay, Paul (2000) *In Praise of Bureaucracy: Weber, organization, ethics*. London: Sage Publications.

Goode, Luke (2005) *Jürgen Habermas: Democracy and the Public Sphere*. London: Pluto Press.

Guba, Egon G. & Lincoln, Yvonna S. (1989) *Fourth Generation Evaluation*. London: Sage Publications.

Habermas, Jürgen (1982) “Reply to my critics”, in Thompson, J. & Held, D. (Eds.), *Habermas: Critical Debates*. Basingstoke: Macmillan.

Harvey, Lee & Newton, Jethro (2004) “Transforming Quality Evaluation”, in *Quality in Higher Education*. Vol. 10, No. 2, pp.149-165. Carfax Publishing.

Holkeri, Katju & Summa, Hilkka (1996) *Contemporary Developments in Performance Management: Evaluation of public management reforms in Finland: from ad hoc studies to a programmatic approach*. Paris: OECD.

Hoppers, Catherine A. Odora (2006) “Evaluation undressed: Citizen’s Demand for Transparency and Good Faith in the Context of Accountability and Democracy”, in Halvorsen, Torill Aagot (ed.) (2006) *Evaluation of Education – on Whose Terms?* Oslo: University of Oslo.

House, Ernest R. (1981) “Assumptions Underlying Evaluation Models”, in Franke-Wikberg, Sigbrit & Lundgren, Ulf P. (Eds.) *Att värdera utbildning* [Evaluating Education]. Stockholm: Vahlström & Widstrand.

Itkonen, Tiina & Jahnukainen, Markku (2007) “An Analysis of Accountability Policies in Finland and the United States”, in *International Journal of Disability, Development and Education,* Vol. 54, No. 1, pp 5-23. Routledge.

Karlsen, Gustav E. (2002) *Utdanning, styring og marked: Norsk utdanningspolitikk i et internasjonalt perspektiv* [Education, governance and market: Norwegian educational policies in an international perspective] Oslo: Universitetsforlaget.

Knežević , Božana (2005) *Programme Evaluation in Higher Education. Theoretical Reflections and Practical Experiences*. Frankfurt am Main: Peter Lang GmbH.

Kohonen, Viljo (1996) *Learning contents and processes in context: towards coherence in educational outcomes through teacher development*, in Niemi, Hannele & Tirri, Kirsi (Eds.) *Effectiveness of Teacher Education. New Challenges and Approaches to Evaluation*. Department of Teacher Education. Tampere University.

Kvale, Steinar (1996) *InterViews: An Introduction to Qualitative Research Interviewing*. Sage Publications.

Lay, Margaret & Papadopoulos, Irena (2007) "An Exploration of Fourth Generation Evaluation in Practice", in *Evaluation*. Vol. 13, No. 4, pp. 495-504. London: Sage Publications.

Layder, Derek (1994) *Understanding Social Theory*. London: Sage Publishing.

Lianos, Michalis (2003) "Social Control after Foucault", in *Surveillance and Society*, Vol. 1, No. 3, pp. 412-430.

Lyytinen, Heikki K. (2006) *New Directions in Educational Evaluation in Finland: the Finnish Education Evaluation Council*, presentation: http://www.hm.ee/index.php?popup=download&id=4430.

Maxwell, Joseph A. (1996) *Qualitative Research Design: An interactive approach*. London: Sage Publications.

McTaggart, R. (1991) "When Democratic Evaluation Doesn't Seem Democratic", in *Evaluation Practice*, Vol. 12, No. 1, pp. 9-21.

Mercer, Malcolm (2005) "Evaluating education projects and programmes: on whose terms and to what effect?" in Halvorsen, Torill Aagot (ed.) (2006) *Evaluation of Education – on Whose Terms?* Oslo: University of Oslo.

Møller, Jorunn; Sivesind, Kirsten; Skedsmo, Guri & Aas, Marit (2006) *Skolelederundersøkelsen 2005: Om arbeidsforhold, evalueringspraksis og ledelse i skolen* [School Leadership Survey 2005: Working Conditions, Evaluation Procedures and Management in School]. Acta Didactica. Department of Teacher Education and School Development. University of Oslo.

Mølstad, Christina (2008) *Teacher Education in Norway and Finland compared: What can Norway learn from the Finnish teacher education about curriculum, recruitment and attitude to globalization?* Oslo: University of Oslo

National Committee for Research Ethics in Social Sciences and the Humanities (2006) *Forskningsetiske retningslinjer for samfunnsvitenskap, humaniora, juss og teologi* [Research ethical guidelines for social sciences, the humanities, law and theology]

Neave, Guy (1998) "The Evaluative State Reconsidered", in *European Journal of Education*. Vol. 33, No. 3, pp 265-283. Paris: European Institute of Education and Social Policy.

Niemi, Hannele (2006) "Den finska forskningsorienterade lärarutbildingen" [The Finnish research based teacher education], in Skagen, Kaare (Ed.) *Lærerutdannelsen i Norden* [Teacher education in the North.] Norwegian Academic Press.

Nordahl, Thomas (2000) *Samarbeid mellom skole og hjem – en kartleggingsundersøkelse*. [Cooperation between school and home – a survey research]. Oslo, Norway: NOVA.

NOU 2002:10 (2002) *Førsteklasses fra første klasse*. [Norwegian Public Report: *First Rate from First Grade*] The Norwegian Ministry of Education and Research.

PISA (2003) *Learning for Tomorrow's World – First Results from PISA 2003.* Paris: OECD.

Pollitt, Christopher & Bouckaert, Geert (2004) *Public Management Reform: A Comparative Analysis*. Oxford University Press.

Power, Michael (1997) *The Audit Society*. Oxford: Oxford University Press.

Reynolds, David & Teddlie, Charles (2001) "Reflections on the Critics, and Beyond Them" in *Schools Effectiveness and School Improvement*. Vol. 12, No. 1, pp. 99-113. Swets & Zeitlinger Publishers.

Sahlberg, Pasi (2007) "Education policies for raising student learning: the Finnish approach", in *Journal of Education Policy*, Vol. 22, No. 2, pp. 147-171.

Schmidtlein, Frank A. (2004) "Assumptions Commonly Underlying Government Quality Assessment Practices", in *Tertiary Education and Management*. Vol. 10, No. 4, pp 263-285. Kluwer Academic Publishers.

Townley, Barbara (1993) "Foucault, Power/Knowledge, and its Relevance for Human Resource Management", *Academy of Management Review*, Vol. 18, No. 3, pp. 518-545.

Trow, Martin (1996) "Trust, markets and accountability in higher education: a comparative perspective", in *Higher Education Policy*, Vol. 9, No. 4, pp 309-324. Elsevier Science Ltd.

Turner, J. (1988) *A Theory of Social Interaction*. Cambridge: Polity Press.

Vogt, Paul W. (1993) *Dictionary of Statistics and Methodology: A nontechnical guide for the social sciences*. Sage Publications.

Webb, R.; Vulliamy, G.; Hakkinen, K. & Hamalainen, K. (1998) "External Inspection or School Self-Evaluation? A Comparison Analysis of Policy and Practice in Primary Schools in England and Finland", in *British Educational Research Journal*. Vol. 25, No. 5, pp 539-56. London: Taylor & Francis, Ltd.

Webb, Rosemary; Vulliamy, Graham; Hämäläinen, Seppo; Sarja, Anneli; Kimonen, Eija & Nevalainen, Raimo (2004) "A comparative analysis of primary teacher professionalism in England and Finland", in *Comparative Education*, Vol. 40, No. 1, pp. 83-107.

Weber, Max (1990) *Makt og byråkrati: Sentrale verker* [*Power and Bureaucracy: Central Works*]. Oslo: Gyldendal.

Weiss, Carol (1986) "The Stakeholder Approach to Evaluation: Origins and Promise", in House, Ernest R. (ed.) *New Directions in Educational Evaluation*. London: The Falmer Press.

Weiss, Carol (1998) *Evaluation*. Prentice Hall.

Westbury, Ian; Hansén, Sven-Erik; Kansanen, Pertti & Björkvist, Ole (2005) "Teacher Education for Research-based Practice in Expanded Roles: Finland's Experience", in *Scandinavian Journal of Educational Research*. Vo. 49, No. 5, pp. 475-485. Routledge.

Yin, Robert K.; Kaftarian, Shakeh Jackie & Jacobs, Nancy F. (1995) "Empowerment Evaluation at Federal and Local Levels", in Fetterman, David M.; Kaftarian, Shakeh & Wandersman, Abraham (Eds.) *Empowerment Evaluation: Knowledge and Tools for Self-Assessment & Accountability*. London: Sage Publications.

Yin, Robert K. (2003) *Case Study Research: Design and Methods*. London: Sage Publications.

Appendices

These appendices present my interview guides for the field work. These worked solely as guiding principles, and nearly none of my interviews featured only these questions. However, this is included to give an impression of what the interview was like.

Appendix 1: Interview guide for principals

1. How long have you had this position?
2. How do you find working in this position?
3. What do you think of the evaluation tools the school is provided with?
4. What do you do with the evaluations?
5. What is your procedure for implementing change in the school based on evaluations?
6. How satisfactory is this process?
7. How does the school as an organization deal with evaluations?
8. To what degree would an external evaluation of the school as organization be of use to the school?
9. How should the division of labor in the school be done?

Appendix 2: Interview guide for teachers

1. For how long have you been working at the school?
2. How satisfied are you with working at the school?
3. What do you know about the evaluations the school goes through every year?
4. What does the school get out of the evaluations it participates in?
5. How does the evaluation process at the school function?
6. What do the evaluations lead to?
7. What do you think of the school management's handling of the feedback evaluations provide?
8. To what degree do the evaluations suit the school sufficiently?
9. How would you have conducted the evaluations?
10. What possibility do you get to reflect around your role as a teacher using these evaluations?